



2015 개정 교육과정 진로 선택 과목

인공지능 수학

CONTENTS

제 1 부

수학과 교육과정 설명 및 자료 개발 방향

I 수학과 교육과정의 주요 개정 내용	6
II 수학과 교육과정의 교수·학습 및 평가	14
III 교수·학습 및 평가 자료 개발 방향	23

제 2 부

교수·학습 및 평가 자료

I 인공지능과 수학	
1. 영역의 개관	30
2. 모듈① 인공지능 발전 과정에서 수학이 어떻게 활용되었을까?	32

II 자료의 표현

1. 영역의 개관	70
2. 모듈② 텍스트 자료는 어떻게 표현할까?	74
3. 모듈③ 이미지 자료는 어떻게 표현하고 처리할까?	98

III 분류와 예측

1. 영역의 개관	116
2. 모듈④ 텍스트 자료는 어떻게 분류할까?	120
3. 모듈⑤ 이미지 자료는 어떻게 분류할까?	138
4. 모듈⑥ 확률을 이용한 예측은 어떻게 할까?	146
5. 모듈⑦ 추세를 이용한 예측은 어떻게 할까?	164

IV 최적화

1. 영역의 개관	174
2. 모듈⑧ 예측의 정확도를 어떻게 높일까?	178



2015 개정 교육과정
교수·학습 자료

인공지능 수학

1

제 부

수학과 교육과정 설명 및 자료 개발 방향

수학과 교육과정의 주요 개정 내용¹⁾

교육과정은 학교 교육이 어떤 방향을 추구해야 하며 무엇을 어떻게 다루어야 하는지를 나타내는 문서이다. 교육부는 인문학적 상상력과 과학 기술 창조력을 갖춘 창의·융합형 인재를 길러내기 위한 교육개혁의 일환으로 2015 개정 교육과정 개발을 추진하였다. 이에 따라 2015 개정 수학과 교육과정은 창의·융합형 인재를 양성하기 위해 교과 역량을 강조하며 교육 내용의 질적인 감축을 이루고 수학 학습에서 정의적 측면을 강조하는 한편 공학적 도구의 활용을 강조하면서, 문·이과 통합형 교육과정의 의의를 살리고 수학 교육의 국제 동향을 반영한 결과로 개발되었다.

01

〈인공지능 수학〉 내용 구성

진로 선택 과목인 〈인공지능 수학〉은 공통 과목인 〈수학〉을 학습한 후, 인공지능 분야에서 수학이 어떻게 활용되는지 알고자 하는 학생들이 선택할 수 있는 과목이다. 〈인공지능 수학〉의 내용은 ‘인공지능과 수학’, ‘자료의 표현’, ‘분류와 예측’, ‘최적화’의 4개 핵심 개념으로 구성된다. ‘인공지능과 수학’에서는 인공지능과 관련된 수학, ‘자료의 표현’에서는 텍스트 자료의 표현과 이미지 자료의 표현, ‘분류와 예측’에서는 자료의 분류와 경향성을 이용한 예측, ‘최적화’에서는 최적화와 의사 결정을 다룬다.

〈인공지능 수학〉의 내용 체계는 **표1**과 같다. 핵심 개념은 수학이라는 학문의 기초적인 개념이나 원리를 말하는 것으로 학습한 내용의 세부 사항을 잊은 후에도 지속되어야 하는 큰 개념을 말하고, 일반화된 지식은 학생들이 알아야 할 내용을 명제적 지식의 형태로 진술한 것이다.

표1 〈인공지능 수학〉의 내용 체계

영역/핵심 개념	일반화된 지식	내용 요소	관련 학습 요소 ²⁾
인공지능과 수학	수학은 인공지능의 발전을 이끌어 왔으며, 인공지능 기술 전반에 활용되고 있다.	· 인공지능과 관련된 수학	· 진리표 · 순서도

1) 1장의 내용은 박경미 외(2015)의 연구 보고서 일부를 요약한 내용임.

자료의 표현	수와 수학 기호는 자료를 효과적으로 표현할 수 있는 도구이고, 인공지능이 다루는 자료는 수학을 이용하여 표현된다.	· 텍스트 자료의 표현 · 이미지 자료의 표현	· 벡터 · 행렬
분류와 예측	인공지능을 이용하면 자료를 정리, 분석하고 패턴을 찾아 새로운 대상에 대한 분류와 예측을 할 수 있으며, 이때 확률과 함수 등이 활용된다.	· 자료의 분류 · 경향성과 예측	· 유사도 · 추세선 · 조건부확률
최적화	인공지능은 자료를 기반으로 합리적인 의사 결정을 내리는 기술을 제공하고, 이때 주어진 자료에 가장 적합한 의사 결정 모델을 찾기 위해 함수를 만들고 최적화하여 문제를 해결한다.	· 최적화와 의사 결정	· 함수의 극한 · 이차함수의 미분계수 · 손실함수 · 경사하강법

02

수학 교과 역량의 구현

2015 개정 교육과정의 가장 큰 특징은 핵심역량(core competency)의 강조이다. 교육과정 총론은 ‘자기 관리 역량’, ‘지식 정보 처리 역량’, ‘창의적 사고 역량’, ‘심미적 감성 역량’, ‘의사소통 역량’, ‘공동체 역량’의 여섯 가지 핵심역량을 도출하였다. 수학과는 교과 특성을 가미하여 2009 개정 교육과정에 명시된 ‘문제 해결’, ‘추론’, ‘의사소통’의 세 가지 수학적 과정에 ‘창의·융합’, ‘정보 처리’, ‘태도 및 실천’의 세 가지를 추가하여 수학 교과 역량을 여섯 가지로 규정하였다. 각각의 교과 역량은 하위 요소와 기능을 갖는데, 하위 요소는 구체적인 역량이고, 기능은 하위 요소를 구현하는 것이다. 기능은 수학 교과 고유의 탐구 기능 및 사고방식을 반영하는 것으로, 수학 내용을 학습한 후 학생들이 할 수 있거나 할 수 있기를 기대하는 도달점 또는 수행 능력을 말한다.

수학 교과 역량은 교육과정의 성격, 목표, 내용 체계, 성취기준, 교수·학습 방법 및 유의 사항, 교수·학습 및 평가 방향에 적극 반영되어 있다. 수학 교과 역량에 대한 설명은 다음과 같다.

2) 인공지능에서 수학이 활용될 때 관련되는 주요 수학적 개념이나 원리를 제시한 것으로, 이러한 수학적 개념이나 원리를 자세히 다루기보다는 인공지능에서 어떻게 활용되는지를 중심으로 다루도록 유의함.

가 문제 해결

문제 해결 능력은 ‘해결 방법을 알고 있지 않은 문제 상황에서 수학의 지식과 기능을 활용하여 해결 전략을 탐색하고 최적의 해결 방안을 선택하여 주어진 문제를 해결하는 능력’을 의미한다. 문제 해결 능력은 수학을 실생활이나 타 교과에 활용하는 도구적 측면뿐만 아니라 수학적 사고의 함양, 수학적 탐구 방법의 습득을 위해서 매우 강조되는 수학과와 핵심역량이라고 할 수 있다. 2015 개정 수학과 교육과정에서는 Polya(1945)의 문제 해결 단계를 반영한 ‘문제 이해 및 전략 탐색’과 ‘계획 실행 및 반성’, 국제 학업 성취도 평가(PISA: Programme for International Student Assessment)의 한 영역인 ‘협력적 문제 해결’, 여러 나라의 교육과정에서 강조되고 있는 ‘수학적 모델링’, 2007 개정 수학과 교육과정부터 문제 해결의 교수·학습 방법으로 강조된 ‘문제 만들기’를 문제 해결 능력의 하위 요소로 두었다.

표 2 문제 해결 능력의 하위 요소와 그 의미 및 기능 (박경미 외, 2015, p.39)

하위 요소	의미	기능
문제 이해 및 전략 탐색	문제에서 구하고자 하는 것과 주어진 조건 및 정보를 파악하고, 적절한 해결 전략을 탐색하여 풀이 계획을 수립하는 능력	(문제) 이해하기, 분석하기, (조건, 정보) 파악하기, (관계) 파악하기, 계획하기, 탐구하기, 일반화하기, 특수화하기, 유추하기, 분류하기, 조사하기, 거꾸로 생각하기, 단순화하기, 그림으로 나타내기, 표 만들기, 식 세우기, (다양한 전략) 구사하기
계획 실행 및 반성	계획한 풀이 과정을 수행하고 검증 및 반성을 통하여 해결 방법과 해답을 평가하는 능력	계산하기, (절차) 수행하기, 문제 해결하기, 적용하기, 활용하기, 점검하기, 반성하기, 평가하기
협력적 문제 해결	균형 있는 책임 분담과 상호작용을 통해 집단적으로 문제 해결을 수행하는 능력	설명하기, 정당화하기, 질문하기, 비판하기, (의견) 존중하기, (의견) 조정하기, 의사결정하기, 토론하기, 제안하기, 종합하기
수학적 모델링	실생활 문제 상황을 수학적으로 나타내고 분석하여 결론을 도출하고 이를 상황에 맞게 해석하는 능력	(상황) 모델링하기, 변환하기, 분석하기, 적용하기, 활용하기, 해석하기, 결론 도출하기, 점검하기
문제 만들기	주어진 문제를 변형하거나 새로운 문제를 만들어 해결하는 능력	(조건) 변형하기, 유사성 찾기, 비교하기, 관련짓기, 확장하기, 생성하기, (문제) 만들기

나 추론

추론 능력은 ‘수학적 사실을 추측하고 논리적으로 분석하고 정당화하며 그 과정을 반성하는 능력’을 의미한다. 수학은 논리적이고 합리적인 사고력을 신장시키는 교과로서, 추론은 수학 교과에서 매우 핵심적이고 중요한 역량이라 할 수 있다. 2015 개정 수학과 교육과정에서는 수학적 사실을 발견하기 위한 ‘관찰과 추측’, 논리적 사고 결과를 나타내는 ‘논리적 절차 수행’, 2009 개정 교육과정부터 추론의 교수·학습 방법에 제시된 ‘수학적 사실 분석’, ‘정당화’, ‘추론 과정의 반성’을 추론 능력의 하위 요소로 두었다.

표 3 추론 능력의 하위 요소와 그 의미 및 기능 (박경미 외, 2015, p.40)

하위 요소	의미	기능
관찰과 추측	관찰과 탐구 상황에서 귀납, 유추 등의 개연적 추론을 사용하여 수학적 사실을 추측하는 능력	관찰하기, 추측하기, 규칙 찾기, 탐구하기, 일반화하기, 특수화하기, 유추하기
논리적 절차 수행	수학적 절차와 수학적 사실 도출 과정을 논리적으로 수행하는 능력	형식화하기, 작도하기, 순서 짓기, 대입하기, 단순화하기, 계산하기, 절차 따르기, 풀기, (해) 구하기, 함수 구하기
수학적 사실 분석	수학적 개념, 원리, 법칙을 분석하는 능력	이해하기, (조건, 정보 등) 파악하기, 분석하기, 정의하기, 관계 짓기, 비교하기, 구별하기, 측정하기, (단위, 식) 변환하기, 공식 유도하기, (수, 개수, 경우의 수) 세기, 어렵하기, 분해하기, 합성하기
정당화	수학적 사실이 참임을 보이기 위해 증거를 제시하고 이유를 설명하는 능력	정당화하기, 반례 찾기, 예증하기, 증명하기, 설명하기, 규칙 정하기
추론 과정의 반성	자신의 추론 과정이 옳은지 비판적으로 평가하고 되돌아보는 능력	반성하기, 되돌아보기, 비판하기, 평가하기, 검토하기, 판단하기, 판별하기, 확인하기

다 창의·융합

창의·융합 능력은 ‘수학의 지식과 기능을 토대로 새롭고 의미 있는 아이디어를 다양하고 풍부하게 산출하고 정교화하며, 여러 수학적 지식, 기능, 경험을 연결하거나 수학과 타 교과나 실생활의 지식, 기능, 경험을 연결·융합하여 새로운 지식, 기능, 경험을 생성하고 문제를 해결하는 능력’을 의미한다. 창의·융합형 인재를 양성하기 위한 2015 개정 교육과정 총론의 방향이 수학 교과에도 반영되어 창의·융합 능력이 교과 역량으로 선정된 것이다. 2015 개정 수학과 교육과정에서는 창의적 사고 기능으로 손꼽히는 ‘독창성’, ‘유창성’, ‘융통성’, ‘정교성’을 창의와 관련된 요소로 두고, 수학의 연결성을 융합의 개념으로 해석하여 ‘수학 내적 연결’, ‘수학 외적 연결 및 융합’을 하위 요소로 두었다.

표 4 창의·융합 능력의 하위 요소와 그 의미 및 기능(박경미 외, 2015, p.41)

하위 요소	의미	기능
독창성	문제 상황에서 새로운 아이디어, 해결 전략, 해결 방법을 찾아내거나 새로운 관점에서 문제를 제기하는 능력	(새로운 관점에서 문제 해결 방법이나 전략) 찾아내기, (새로운 관점에서) 문제 제기하기, 발견하기, 창작하기, 상상하기, 발명하기, 만들기
유창성	문제 상황에서 많은 아이디어나 해결 방법, 해답을 산출하는 능력	(많은 해결 방법이나 해답) 찾아보기, (문제 해결 방법이나 전략을 2개 이상) 제시하기, (개방형 문제에서 다양한 해답) 산출하기
융통성	고정된 사고방식에서 벗어나 다양한 관점에서 해결 방법이나 전략, 아이디어를 찾아내거나 문제를 제기하는 능력	(다양한 관점에서 해결 방법이나 전략, 아이디어) 찾아내기, 여러 범주(대수, 기하, 식, 표, 그래프 등)에서 해결책 찾아내기, (다양한 관점에서) 문제 제기하기
정교성	기존의 수학적 아이디어에 세부 사항을 추가하거나 변형하여 더욱 가치 있는 것으로 발전시키는 능력	(수학적 아이디어) 구체화하기, (수학적 사실을 표, 그림, 모델, 수학 용어, 기호 등을 사용하여) 간단명료하게 표현하기, (수학적 아이디어나 문제 풀이 과정) 정련하기/정교화하기, (여러 풀이나 설명 중에서) 완결성 높은 것 찾아보기
수학 내적 연결	여러 수학적 지식, 기능, 경험 등을 연결하여 새로운 수학적 지식, 기능, 경험 등을 생성하고 수학 문제를 해결하는 능력	(서로 다른 주제 또는 서로 다른 학년의 수학 지식, 기능, 경험 사이의) 관계 찾기/관련짓기/연결하기/통합하기/재구성하기, (수학 문제 상황에 두 가지 이상의 지식, 기능) 적용하기/문제 해결하기
수학 외적 연결 및 융합	수학과 타 교과나 실생활의 지식, 기능, 경험 등을 연결·융합하여 새로운 지식, 기능, 경험 등을 생성하고 문제를 해결하는 능력	(실생활이나 타 교과 상황과 관련된 수학적 지식, 기능, 경험 등) 찾아보기, (실생활이나 타 교과 상황에 수학적 지식, 기능, 경험 등) 적용하기/연결하기/관련짓기/융합하기

라 의사소통

의사소통 능력은 ‘수학 지식이나 아이디어, 수학적 활동의 결과, 문제 해결 과정, 신념과 태도 등을 말이나 글, 그림, 기호로 표현하고 다른 사람의 아이디어를 이해하는 능력’을 의미한다. 수학을 이용하는 능력은 기호 및 수학 용어의 학습을 통해 이루어지며, 이것은 수학적 언어가 자연스럽게 사용되는 상황 속에서 학생들이 읽고 쓰고 아이디어를 토론하는 기회를 가짐으로써 가장 잘 성취될 수 있다(NCTM, 1989). 2015 개정 수학과 교육과정에서는 NCTM(2000) 과정 규준의 하나인 표현을 ‘수학적 표현의 이해’, ‘수학적 표현의 개발 및 변환’으로 구분하여 하위 요소로 선정하였고, 2009 개정 수학과 교육과정의 의사소통 교수·학습 방법에서 수학에 대한 ‘자신의 생각 표현’과 ‘타인의 생각 이해’도 의사소통의 하위 요소로 두고 있다.

표 5 의사소통 능력의 하위 요소와 그 의미 및 기능 (박경미 외, 2015, p.42)

하위 요소	의미	기능
수학적 표현의 이해	수학적 표현의 의미를 이해하고 정확하게 사용하는 능력	그리기, (수, 시각) 읽기·쓰기, 표현하기, 형식화하기, 서술하기, 작도하기, 이해하기
수학적 표현의 개발 및 변환	자신의 아이디어를 나타내는 표현을 만들고 수학적 표현들 끼리 변환하는 능력	(표) 만들기, (그래프) 그리기, 꾸미기, 채우기, 이름짓기, (그림, 식, 표 등으로) 나타내기, 표현하기, 선택하기, 변환하기 (바꾸기)
자신의 생각 표현	수학 학습 활동 과정과 결과를 다른 사람에게 표현하는 능력	설명하기, 쓰기, 말하기, 보여주기, 토론하기
타인의 생각 이해	다른 사람의 생각을 이해하고 평가하는 능력	경청하기, 질문하기, 파악하기, 토론하기

마 정보 처리

정보 처리 능력은 ‘다양한 자료와 정보를 수집·정리·분석·활용하고 적절한 공학적 도구나 교구를 선택·이용하여 자료와 정보를 효과적으로 처리하는 능력’을 의미한다. 정보 처리 능력은 최근 강조되고 있는 정보과학적 사고(Computational Thinking), 지식 정보화 시대를 살아가면서 다양한 정보를 해석하고 비판적으로 평가하며 의사소통하는 통계적 소양과 밀접하게 관련된다. 학생들이 문제를 제안하고 이에 답하기 위하여 자료를 수집, 조직, 표현하며, 탐색적인 방법으로 자료를 해석하여 자료에 근거한 추론, 예측, 논의를 발전시키고 평가하는 것은 학생들의 삶 전체에 유용한 사고 도구와 사고 방법을 제공할 수 있다(NCTM, 1989). 뿐만 아니라 실생활 및 수학적 문제 상황에서 정보를 효과적으로 처리하기 위해 적합한 공학적 도구 및 교구를 선택하고 사용하는 능력은 정보화 사회의 필수 소양이다. 이에 2015 개정 수학과 교육과정에서는 ‘자료와 정보 수집’, ‘자료와 정보 정리 및 분석’, ‘정보 해석 및 활용’, ‘공학적 도구 및 교구 활용’을 정보 처리 능력의 하위 요소로 두었다.

표 6 정보 처리 능력의 하위 요소와 그 의미 및 기능 (박경미 외, 2015, p.42)

하위 요소	의미	기능
자료와 정보 수집	실생활 및 수학적 문제 상황에서 적절한 자료와 정보를 탐색 및 생성하여 수집하는 능력	(자료를) 수집하기, 조사하기, 기록하기, 탐색하기, 생성하기
자료와 정보 정리 및 분석	수집한 자료와 정보를 목적에 맞게 분류, 정리, 분석, 평가하는 능력	표현하기, 분류하기, 정리하기, 열거하기, 배열하기, 비교하기, 묶기, 분석하기, 분류하기, 분할하기, 시각화하기, 평가하기

정보 해석 및 활용	분석한 정보에 내재된 의미를 올바르게 파악하여 해석, 종합, 활용하는 능력	예측하기, 설명하기, 해석하기, 종합하기, 활용하기
공학적 도구 및 교구 활용	수학적 아이디어와 개념을 탐 구하고 문제를 해결하는 데 적 합한 공학적 도구 및 교구를 선 택하고 이용하는 능력	선택하기, 조작하기, 공학적 도구 활용하 기, 시각화하기

바 태도 및 실천

태도 및 실천 능력은 ‘수학의 가치를 인식하고 자주적 수학 학습 태도와 민주 시민 의식을 갖추어 실천하는 능력’을 의미한다. 2015 개정 수학과 교육과정에서는 학습자의 정의적 측면을 강조하기 위해 교과 역량의 하나로 태도 및 실천 능력을 제시하였으며, 수학 교육을 통해서 학생들이 수학에 대해 흥미와 관심을 갖는 것에서 나아가 수학과 수학 학습, 생활에서 민주 시민으로서의 성향과 생활 태도를 가지려는 의지를 지녀야 한다고 보았다. 태도 및 실천 능력의 하위 요소는 ‘가치 인식’, ‘자주적 학습 태도’, ‘시민 의식’이다.

표 7 태도 및 실천 능력의 하위 요소와 그 의미 및 기능 (박경미 외, 2015, p.43)

하위 요소	의미	기능
가치 인식	수학에 대해 관심과 흥미를 갖고, 수학의 실용적, 도야적, 심미적, 문화적 가치를 인식하는 능력	(수학에 대해) 관심과 흥미 갖기, (수학의) 가치 인식하기, (수학의) 역할 이해하기, (수학의) 필요성/유용성 인식하기, (수학의) 편리함 인식하기
자주적 학습 태도	수학 학습 의지와 자신감, 끈기를 갖고 자신 스스로 목표를 설정하여 자율적으로 학습을 수행하며 학습 결과를 평가하는 태도	(즐거움, 성취감, 동기화, 안정감, 만족감, 도전 의식, 적극성, 자신감, 끈기) 갖기, 목표 설정하기, 계획 세우기, 조절하기, 점검하기, 평가하기, 시간 관리하기, 자율적으로 행동하기
시민 의식	수학적 활동을 통하여 정직하고 공정하며 책임감 있게 행동하고 어려움을 극복하기 위해 도전하는 용기 있는 태도, 타인을 배려하고 존중하며 협력하는 태도, 논리적 근거를 토대로 의견을 제시하고 합리적으로 의사결정 하는 태도를 갖고 이를 실천하는 능력	(공정, 정직)한 태도 취하기, (책임감, 도전 정신, 용기) 갖기, 배려하기, 존중하기, 협력하기, 논리적 근거를 토대로 의견 제시하기, 이유 설명하기, 합리적으로 의사결정하기

6개 교과 역량은 <인공지능 수학>에서 ‘교수·학습 방법 및 유의 사항’에 반영되어 있다.

03

학습자의 정의적 측면 강조

국제 학업 성취도 평가 결과 우리나라 학생들의 인지적 측면의 성취도에 비해 정의적 측면의 성취도는 다소 미흡한 것으로 나타나고 있다. 이에 따라 2015 개정 수학과 교육과정에서는 정의적 측면을 태도 및 실천이라는 교과 역량과 교수·학습 방법 및 유의 사항에 반영하여 학교 현장의 실제적인 변화를 도모하였다. ‘교수·학습 방법 및 유의 사항’에서는 각 내용 주제에 대한 필요성, 유용성, 가치 인식을 강조하고 있는데, 다음은 <인공지능 수학> ‘교수·학습 방법 및 유의 사항’의 예이다.

- 실생활에서 활용되는 인공지능에서 수학이 활용되는 다양한 사례를 찾아보는 활동을 하게 한다.
- 인공지능과 관련된 수학에 대한 평가를 할 때는 학생들의 흥미, 자신감, 수학의 가치 인식, 학습 태도 등 정의적 영역의 평가를 할 수 있다.
- 실생활의 텍스트 자료로 댓글이나 감상평 등을 다룰 수 있다.

04

공학적 도구의 활용 강조

수학은 엄밀한 형식 체계를 갖지만 학생들에게 그 체계를 그대로 전달하거나 성급하게 형식화하는 것은 수학을 어렵게 하는 원인이 된다. 공학적 도구의 활용은 이러한 교수·학습을 보완할 수 있는 방안이 될 수 있다. 공학적 도구를 활용하면 추상적인 수학 내용을 시각화할 수 있어 학습자가 수학의 개념, 원리, 법칙을 구체화하여 이해하는 데 도움이 되고, 복잡한 계산이나 대수적인 문자식의 처리를 계산기나 컴퓨터가 신속하게 대행함으로써 사고력 중심의 교수·학습 활동에 전념할 수 있다.

제6차 수학과 교육과정부터 계산기나 컴퓨터 등 공학적 도구를 활용하는 교수·학습 방법이 이미 제시되었으나 일반적인 ‘교수·학습 방법’에서 언급되어 수학 수업 현장에서의 활용은 미흡한 편이었다. <인공지능 수학>에서는 ‘교수·학습 방법 및 유의 사항’에서 교과서와 수업에서 공학적 도구가 적극적으로 활용되어야 함을 명시적으로 강조하고 있다. 따라서 본 인공지능 수학 교과에서는 기존의 수학 교과보다 공학적 도구를 보다 적극적으로 활용할 것을 제안하고 있는바, 알지오매스와 같은 수학 앱을 활용하거나 수학 실험실 등을 활용하여 파이썬의 프로그래밍을 활용하여 개념을 이해하거나 탐구하는 활동을 할 수 있다.

- 텍스트 자료나 이미지 자료를 수와 수학 기호로 표현할 때 공학적 도구를 이용할 수 있다.
- 텍스트 자료를 표, 그림, 그래프, 단어 구름 만들기 등을 이용하여 시각화한 후 이를 해석해 보는 활동을 할 수 있으며, 이때 공학적 도구를 이용할 수 있다.
- 추세선은 y 의 측정값과 예측값 사이의 오차를 구하여 이를 최소화한다는 의미 정도로 간단히 다루고, 추세선을 구할 때 공학적 도구를 이용할 수 있다.

II

수학과 교육과정의 교수·학습 및 평가

01

교수·학습

2015 개정 수학과 교육과정에서 추구하는 수학 교과 역량 함양과 학습 내용 감축은 교수·학습 방법의 개선을 통해 성취될 수 있다. 이를 위해 2015 개정 수학과 교육과정은 교수·학습에 대한 내용을 ‘교수·학습 원칙’과 ‘교수·학습 방법’으로 구분하여 이전 교육과정보다 구체적으로 설명하였다.

가 교수·학습 원칙

교수·학습 원칙에서는 수학 교과의 성격이나 특성에 비추어 수학 학습의 목적과 가치를 제시하여 교사가 학생을 지도할 때 방향성 정립에 도움을 주고자 하였고, 교육과정 개정의 취지에서 강조한 창의·융합형 인재 양성, 학습자 중심 교육, 교과 역량 함양 등을 구현할 수 있도록 수학과 교수·학습의 일반적인 방향과 원칙을 제시하였다(교육부, 2015).

교수·학습 원칙에서 개정의 주요 내용 첫 번째는 수학과 교수·학습의 목적을 제시한 점이다. 수학 교육의 목적은 성격에서 ‘수학 학습을 통해 학생들은 수학의 규칙성과 구조의 아름다움을 음미할 수 있고, 수학의 지식과 기능을 활용하여 수학 문제뿐만 아니라 실생활과 다른 교과의 문제를 창의적으로 해결할 수 있으며, 나아가 세계 공동체의 시민으로서 갖추어야 할 합리적 의사 결정 능력과 민주적 소통 능력을 함양할 수 있다’로 이미 기술되어 있지만, 수업을 진행하는 교사들이 실질적으로 참고하는 교수·학습 부분에서 기억하기 용이하도록 간략하게 진술된 목적이 필요하고 수학 수업을 하기 전에 가져야 할 철학적 기초를 제공하기 위하여 (가) 항에서 목적이 진술되었다.

둘째, 교육과정이 학교 현장에서 실효성을 갖도록 하기 위해 교육 목표, 내용, 교수·학습, 평가의 일관성을 강조하였다. 교사들이 수학과 목표를 추상적으로만 인식하고 교육과정에서 성취기준만을 지도하기보다 교육 목표를 성취하기 위해 교육과정에 제시된 성취기준을 지도하기에 가장 적절한 교수·학습 방법을 선택하고 그에 따라 평가를 실시할 수 있도록 하는 원칙을 (나) 항에 제시하였다. 특히 선행교육 방지를 위한 공교육정상화법이 발효되고 있는 상황에서 수학과 교수·학습에서 교육 내용을 준수하는 내용도 포함시켰다.

셋째, 2015 개정 수학과 교육과정에서 크게 강조하고 있는 교과 역량을 함양시키기 위해 일반적으로 구비되어야 할 여건과 일반적인 원칙을 (다) 항에 제시하였다.

이러한 개정 내용에 따라 교수·학습의 원칙 (가) 항은 수학과 교수·학습의 목적, (나) 항은 교수·학습에서 교육 내용의 준수와 목표, 내용, 평가와의 일관성 유지, (다) 항은 교과 역량을 함양하기 위한 환경 및 교수·학습 운영, (라) 항은 교육 내용의 재구성, (마) 항은 보충·심화 학습에 대한 기회 제공의 내용을 담고 있다. 이 내용은 교사가 수학과 교육과정에 근거하여 수학과 교수·학습 계획을 수립할 때 고려해야 할 원칙이다.

- (가) 수학과 교수·학습은 학생이 수학과 교육과정에 제시된 목표를 달성하고 전인적으로 성장하도록 돕는 것을 목적으로 한다.
- (나) 수학과 교수·학습은 교육과정에 제시된 내용의 수준과 범위를 준수하고, 교육과정에 제시된 목표, 내용, 평가와 일관성을 가져야 한다.
- (다) 문제 해결, 추론, 창의·융합, 의사소통, 정보 처리, 태도 및 실천과 같은 수학 교과 역량을 함양하기 위한 교육 환경을 조성하고, 이에 적합한 교수·학습을 운영한다.
- (라) 과목별 내용의 배열 순서가 반드시 교수·학습의 순서를 의미하는 것은 아니므로 교수·학습 계획을 수립하거나 학습 자료를 개발할 때에는 내용의 특성과 난이도, 학교 여건, 학생의 수준 및 개인차 등을 고려하여 내용, 순서 등을 재구성할 수 있다.
- (마) 교육과정에 제시된 내용을 지도할 때 학습 결손이 있는 학생에게는 보충 학습, 우수 학생에게는 심화 학습의 기회를 추가로 제공할 수 있다.

나 교수·학습 방법

교수·학습 방법에서는 교수·학습 원칙에 부합하는 적절한 방법 및 전략을 제시하였다. 학습자 중심의 학습을 강조하고, 학습자의 능력, 수준, 흥미, 필요, 학습 양식에 따른 교수·학습 방법으로 설명식 교수, 탐구 학습, 프로젝트 학습, 토의·토론 학습, 협력 학습, 매체 및 도구 활용 학습의 특징과 방법을 제시하였다. 또한 수학 교과 역량인 문제 해결, 추론, 창의·융합, 의사소통, 정보 처리, 태도 및 실천 능력을 함양하기 위한 구체적인 교수·학습 방법과 유의점을 제시하였다. 그리고 발문과 수준별 수업 운영 방법에 대한 유의점도 제시하였다. 이를 구체적으로 살펴보면 아래와 같다.

교수·학습 방법 (가) 항에서는 수학과 성격 고려하여 학생들에게 적용하기 적합한 교수·학습 방법을 여섯 가지로 제시하였다. 교수·학습 방법별로 특징을 제시하고, 그 교수·학습 방법을 실시했을 때 기대되는 효과와 유의점을 함께 설명하였다. 2009 개정 수학과 교육과정에서는 발견 학습, 탐구 학습, 협동 학습, 개별 학습, 설명식 교수의 방법이 제안되었으나 2015 개정 수학과 교육과정에서는 최신 교수·학습 방법을 반영하고 그 의미를 보다 구체화하여 교수·학습 방법으로 제시하였다.

(가) 수학과와 수업은 학생의 능력과 수준 등을 고려하여 설명식 교수, 탐구 학습, 프로젝트 학습, 토의·토론 학습, 협력 학습, 매체 및 도구 활용 학습 등을 적절히 선택하여 적용한다.

- ① 설명식 교수는 교사가 설명과 시연을 통해 수업을 주도하는 교수·학습 방법으로, 수업 내용을 구조화하여 체계적으로 지도하는 데 효과적이다. 이때, 교사는 학생의 적극적인 수업 참여를 유도하고, 사고를 촉진하는 발문을 적절히 활용한다.
- ② 탐구 학습은 학생이 중심이 되어 수학 개념, 원리, 법칙을 발견하고 구성하는 교수·학습 방법으로, 학생 스스로 자료와 정보로부터 지식을 도출하거나 지식의 타당성을 확인하는 능력을 기를 수 있게 한다.
- ③ 프로젝트 학습은 특정 주제나 과제를 탐구하기 위해 계획을 수립하고 수행하여 결과물을 산출하거나 발표하는 교수·학습 방법으로, 개인별 또는 집단별로 실시할 수 있다.
- ④ 토의·토론 학습은 특정 주제에 대해 협의하거나 논의하는 교수·학습 방법으로, 의사소통이 지니는 상호 협력적인 면을 강조한다. 이를 통해 학생들이 교과 내용을 폭넓게 이해하고 논리적이고 비판적으로 추론하며, 다른 사람의 의견을 비판적으로 수용하고, 자신의 주장을 효과적으로 표현하는 능력을 기를 수 있게 한다.
- ⑤ 협력 학습은 모둠 내의 상호작용, 의사소통, 참여를 통해 공동의 학습 목표에 도달하도록 하는 교수·학습 방법으로, 다른 사람을 존중하고 배려하며 모둠 내의 역할을 이해하고 책임감을 기를 수 있게 한다.
- ⑥ 매체 및 도구 활용 학습은 학생의 수준과 학습 내용에 적합한 매체와 도구를 활용하여 흥미를 유발하고 학습의 효율성과 다양성을 도모하는 교수·학습 방법으로, 시청각 자료, 멀티미디어나 인터넷 등의 컴퓨터 활용 매체와 교구, 계산기, 교육용 소프트웨어 등의 도구를 이용한다.

교수·학습 방법 (나) 항부터 (사) 항까지는 수학 교과 역량 각각에 대한 교수·학습 방법을 제안하였다. 2009 개정 수학과 교육과정에서도 수학적 창의력, 문제 해결력, 추론 능력, 의사소통 능력, 인성 함양과 긍정적 태도 신장에 대한 교수·학습 방법을 제안하였으나, 2015 개정 수학과 교육과정에서는 강조된 교과 역량이 교수·학습 방법으로 구현될 수 있도록 그 구체적인 방안을 제안하였다. 교과 역량별로 강조되는 사항은 교과 역량의 하위 요소를 중심으로 구성되었다.

(나) 문제 해결 능력을 함양하기 위한 교수·학습에서는 다음 사항을 강조한다.

- ① 인공지능에 활용되는 수학 사례와 관련된 문제를 해결할 때에는 문제를 이해하고 해결 전략을 탐색하며 해결 과정을 실행하고 검증 및 반성하는 단계를 거치도록 한다.

- ② 협력적 문제 해결 과제에서는 균형 있는 책임 분담과 상호작용을 통해 동료들과 협력하여 문제를 해결하게 한다.
- ③ 수학적 모델링 능력을 신장하기 위해 생활 주변이나 사회 및 자연 현상 등 다양한 맥락에서 인공지능에 활용되는 수학 사례를 찾고 관련 수학적 개념, 원리, 법칙을 탐구하고 이를 일반화하게 한다.
- ④ 문제 해결력을 높이기 위해 인공지능에 활용되는 수학 사례의 문제를 변형하거나 새로운 문제를 만들어 해결하고 그 과정을 검증하는 문제 만들기 활동을 장려한다.

(다) 추론 능력을 함양하기 위한 교수·학습에서는 다음 사항을 강조한다.

- ① 관찰과 탐구 상황에서 귀납, 유추 등의 개연적 추론을 사용하여 학생 스스로 인공지능에 활용되는 수학 사례에서 수학적 사실을 추측하고 적절한 근거에 기초하여 이를 정당화할 수 있게 한다.
- ② 인공지능에 활용되는 수학 사례의 원리, 법칙을 도출하는 과정과 수학적 절차를 논리적으로 수행하게 한다.
- ③ 추론 과정이 맞는지 비판적으로 평가하고 반성하도록 한다.

(라) 창의·융합 능력을 함양하기 위한 교수·학습에서는 다음 사항을 강조한다.

- ① 새롭고 의미 있는 아이디어를 다양하고 풍부하게 산출할 수 있는 수학적 과제를 제공하여 학생의 창의적 사고를 촉진시킨다.
- ② 인공지능에 활용되는 수학 사례와 관련된 하나의 문제를 여러 가지 방법으로 해결하는 예를 제시한 후, 해결 방법을 비교하여 더 효율적인 방법을 찾거나 정교화하게 한다.
- ③ 인공지능에 활용되는 수학의 사례로부터 수학적 지식, 기능, 경험을 연결하거나 수학과 타 교과 및 실생활의 지식, 기능, 경험을 연결·융합하여 새로운 지식, 기능, 경험을 생성하고 문제를 해결하게 한다.

(마) 의사소통 능력을 함양하기 위한 교수·학습에서는 다음 사항을 강조한다.

- ① 수학 용어, 기호, 표, 그래프 등의 수학적 표현을 이해하고 정확하게 사용하며, 수학적 표현을 만들거나 변환하는 활동을 하게 한다.
- ② 인공지능에 활용되는 수학 사례에서 수학적 아이디어 또는 수학 학습 과정과 결과를 말, 글, 그림, 기호, 표, 그래프 등을 사용하여 다른 사람과 효율적으로 의사소통할 수 있게 한다.
- ③ 다양한 관점을 존중하면서 다른 사람의 생각을 이해하고 수학적 아이디어를 표현하며 토론하게 한다.

(바) 정보 처리 능력을 함양하기 위한 교수·학습에서는 다음 사항을 강조한다.

- ① 실생활 및 수학적 문제 상황에서 인공지능 수학과 관련된 자료를 탐색하여 수집하고, 목적에 맞게 정리, 분석, 평가하며, 분석한 정보를 문제 상황에 적합하게 활용할 수 있게 한다.

- ② 교수·학습 과정에서 적절한 교구를 활용한 조작 및 탐구 활동을 통해 수학의 개념과 원리를 이해하도록 한다.
- ③ 계산 능력 배양을 목표로 하지 않는 교수·학습 상황에서는 수학의 개념, 원리, 법칙의 이해와 문제 해결에 초점을 두기 위해 계산기, 컴퓨터, 교육용 소프트웨어 등의 공학적 도구를 이용할 수 있게 한다.

(사) 태도 및 실천 능력을 함양하기 위한 교수·학습에서는 다음 사항을 강조한다.

- ① 수학을 생활 주변과 사회 및 자연 현상과 관련지어 지도하여 수학의 필요성과 유용성을 알게 하고, 수학의 역할과 가치를 인식할 수 있게 한다.
- ② 인공지능 수학에 대한 관심과 흥미, 호기심과 자신감을 갖고 수학 학습에 적극적으로 참여하게 하며, 끈기 있게 도전하도록 격려하여 학습 동기를 유발한다.
- ③ 학생 스스로 목표를 설정하고 학습을 수행하며 학습 결과를 평가하는 자주적 학습 습관과 태도를 갖게 한다.
- ④ 수학적 활동을 통하여 정직하고 공정하며 책임감 있는 태도, 어려움을 극복하기 위해 도전하는 용기 있는 태도, 타인을 배려하고 존중하며 협력하는 태도, 논리적 근거를 토대로 의견을 제시하고 합리적으로 의사를 결정하는 태도를 갖고 이를 실천하게 한다.

교수·학습 방법 (아) 항은 교사와 학생 간의 상호작용에서 중요한 발문에 대한 유의 사항을 제안하였다. 제5차 수학과 교육과정부터 발문의 역할이 언급되었으나 2015 개정 수학과 교육과정에서는 학습자 중심으로 상호작용을 유발하기 위한 구체적인 교수·학습 방법을 제안하였다.

(아) 의미 있는 발문을 하기 위하여 교수·학습에서 다음 사항에 유의한다.

- ① 학생의 사고를 촉진하는 다양한 발문을 통해 상호작용이 활발한 교실 환경을 구축하고 학생의 능동적 수업 참여를 독려한다.
- ② 학생의 인지 발달과 경험을 고려하여 발문하고, 학생의 반응을 의미 있게 처리한다.

교수·학습 방법 (자) 항은 수학 수업에서 수준별 수업을 운영할 때의 유의 사항을 설명하였다. 2009 개정 수학과 교육과정의 교수·학습 방법에 제시된 수준별 수업을 교실 내 소집단이나 수준별 학급으로 편성할 수 있음을 추가하여 설명하였다.

(자) 개인차를 고려하여 수준별 수업을 운영할 때에는 다음 사항에 유의한다.

- ① 학습 목표를 효과적으로 달성하기 위해 교실 내에서 개인차를 고려한 소집단을 구성하거나 수준별 학급을 구성하여 교수·학습을 전개한다.

- ② 수준별 수업을 위해 집단을 편성할 때에는 학생 개인의 능력과 수준, 적성과 희망 등의 학생 상황과 교사 수급과 유휴 교실 등의 학교 상황을 고려한다.
- ③ 수준별 수업은 내용 요소를 차별화하기보다는 내용의 깊이나 접근 방법에 차이를 두어 진행한다.

교수·학습 방법 (차) 항은 수학 수업에서 공학적 도구를 활용한 수업을 운영할 때의 유의 사항을 설명하였다. 2015 개정 수학과 교육과정의 교수·학습 방법에 제시된 공학적 도구를 활용할 시 유의점에 웹 기반 소프트웨어를 사용하여 텍스트코딩을 다룰 때의 유의점을 추가하여 설명하였다.

(차) 공학적 도구를 활용한 수업을 운영할 때에는 다음 사항에 유의한다.

- ① 공학적 도구를 활용하여 인공지능 기술을 직접 시연해 볼 수 있는 환경을 제공함으로써 인공지능에 활용되는 수학을 경험할 수 있는 기회를 제공한다.
- ② 공학적 도구로 PC나 모바일 기기를 활용할 수 있으며, 설치형 또는 웹 기반의 소프트웨어를 사용하여 텍스트코딩을 다룰 수 있다.
- ③ 텍스트코딩을 다룰 때는 학생들이 직관적으로 이해할 수 있는 수준으로 작성된 프로그램의 코드를 제공하여 프로그래밍에 대한 부담을 느끼지 않게 하고, 제공된 프로그램 코드의 숫자를 학생이 부분적으로 수정하면서 자연스럽게 활용할 수 있도록 한다.

02 평가

2015 개정 수학과 교육과정에서는 평가 개선에 실질적인 도움이 될 수 있도록 이전 수학과 교육과정에서 다소 추상적으로 기술된 평가 관련 내용을 보다 구체적으로 기술하였다. 이를 위해 평가 방향을 ‘평가 원칙’과 ‘평가 방법’으로 구분하여 구체적인 설명을 제시하였다.

가 평가 원칙

평가 원칙에서는 수학과 평가의 일반적인 방향과 원칙을 제시하였다.

평가 원칙에서 개정의 주요 내용 첫 번째는 교육 목표, 내용, 교수·학습 방법과 평가의 일관성을 제시한 점이다. 수학 교육의 개선에 있어 강조되는 교수·학습 방법의 변화는 교수·학습 방법의 변화에 부응하는 평가가 일관되게 시행될 때 그 효과를 누릴 수 있다. 수학과 목표, 내용, 교수·학습 방법과 평가의 일관성은 교수·학습 원칙에서 이미 언급되었으나 평가 원칙 (나) 항에서 재차 강조하였다.

둘째, 2015 개정 수학과 교육과정에서 설정한 수학 교과 역량에 대한 평가 방향을 제시하였다. 여러 교과 역량을 평가하되 특정 역량에 치우치지 않도록 균형 있는 평가가 이루어져야 함을 평가 원칙 (다) 항에 명시하였다.

셋째, 다양한 평가 방법을 활용해야 학생들의 수학 학습과 역량에 대한 종합적인 평가가 이루어질 수 있다. 기존의 교육과정도 이미 다양한 평가 방법을 언급하였으나 2015 개정 수학과 교육과정에서는 (라) 항에서 학생의 수준, 평가 목적과 내용을 고려하여 평가 방법을 활용할 것을 구체적으로 제시하였다.

넷째, 평가의 목적이 학생의 수학 학습과 전인적인 성장을 돕고 교사의 교수·학습을 개선하는 것이라면 평가 결과가 의미 있게 활용되어야 한다. 따라서 평가 원칙 (마) 항에서 평가 결과 환류에 대한 내용을 제시하였다.

이러한 개정 내용에 따라 평가 원칙 (가) 항에서는 평가의 목적, (나) 항에서는 교육과정에 제시된 목표, 내용, 교수·학습과의 일관성, (다) 항에서는 수학적 지식과 교과 역량 등 평가의 내용, (라) 항에서는 다양한 평가 방법의 활용, (마) 항에서는 평가 결과의 활용에 대하여 안내하였다.

- (가) 수학과 평가는 학생의 인지적 영역과 정의적 영역에 대한 유용한 정보를 수집·활용하여 학생의 수학 학습과 전인적 성장을 돕고 교사의 수업 방법을 개선하는 것을 목적으로 한다.
- (나) 수학과 평가는 교육과정에 제시된 내용의 수준과 범위를 준수하고, 교육과정에 제시된 목표, 내용, 교수·학습과 일관성을 가져야 한다.
- (다) 수학과 평가에서는 수학의 개념, 원리, 법칙, 기능뿐만 아니라 문제 해결, 추론, 창의·융합, 의사소통, 정보 처리, 태도 및 실천과 같은 수학 교과 역량을 균형 있게 평가한다.
- (라) 수학과 평가는 학습자의 수준을 고려하고 평가 목적과 내용에 따라 다양한 평가 방법을 활용한다.
- (마) 평가 결과는 학생, 학부모, 교사 등에게 환류하여 학생의 수학 학습 개선을 도울 수 있게 한다.

나 평가 방법

평가 방법에서는 학습 과정과 결과에 대한 종합적인 평가, 수업 국면에 따른 진단·형성·총괄평가 실시, 교과 역량 평가와 관련한 다양한 평가 방법, 평가에서 공학적 도구 및 교구 활용에 대한 안내를 하였다.

평가 방법 (가) 항은 학생들의 학습 결과뿐 아니라 그 과정도 평가하도록 하여 교수·학습과 평가가 일관성을 가질 수 있도록 ‘과정 중심 평가’를 실시할 것을 강조하였다.

(가) 수학과 평가는 학습 결과 평가뿐만 아니라 과정 중심 평가도 실시하여 종합적인 수학 학습 평가가 될 수 있게 한다.

평가 방법 (나) 항은 수업 국면에 따라 실시하는 평가의 유형을 진단평가, 형성평가, 총괄평가로 제시하였다. 진단평가는 학생의 선수 학습 내용 확인, 형성평가는 교수·학습의 수정, 총괄평가는 학생의 진전 상태를 확인하는 데 목적이 있다. 교사가 교수·학습 과정에서 지속적인 평가를 통해 학생들이 수학 학습에서 보여주는 다양한 특성을 수집하고 이를 수업에 활용할 수 있어야 한다는 점을 설명하였다.

(나) 수업의 전개 국면에 따라 진단평가, 형성평가, 총괄평가를 적절히 실시하되, 지속적인 평가를 통해 다양한 정보를 수집하고 수업에 활용한다.

평가 방법 (다) 항은 지필 평가, 프로젝트 평가, 포트폴리오 평가, 관찰 평가, 면담 평가, 구술 평가, 자기 평가, 동료 평가 등의 다양한 평가 방법을 수학 교과 역량 평가와 관련지어 제시함으로써 2015 개정 수학과 교육과정의 중점인 수학 교과 역량이 평가에서도 일관성 있게 구현될 수 있도록 하였다. 또한 평가 방법이 각각 어떤 것이며 어떤 능력을 평가하기에 적절하고 어떤 장점이 있는지에 대해 상세히 진술하였고, 양적 평가뿐 아니라 질적 평가도 가능함을 명시하였다.

(다) 학생의 수학 학습 과정과 결과는 포트폴리오 평가, 프로젝트 평가, 지필 평가, 관찰 평가, 면담 평가, 구술 평가, 자기 평가, 동료 평가 등의 다양한 평가 방법을 사용하여 양적 또는 질적으로 평가한다.

① 포트폴리오 평가는 일정 기간 동안 수학 학습 수행과 그 결과물을 평가하는 방법으로, 학생의 학습 내용 이해와 수학 교과 역량을 종합적으로 판단하고 학생의 성장에 대한 정보를 얻는 데 활용할 수 있다.

② 프로젝트 평가는 수학 학습을 토대로 인공지능에 활용되는 수학 사례와 관련된 특정한 주제나 과제에 대해서 자료를 수집하여 분석하고, 종합하여 문제를 해결하는 과정과 결과를 평가하는 방법으로 문제 해결, 창의·융합, 정보 처리 능력 등을 평가할 때 활용할 수 있다. 공학적 도구를 활용하여 실습하고 탐구하고 산출물을 발표할 수 있도록 한다.

③ 지필 평가는 인공지능에 활용되는 수학 사례와 관련된 문제 해결, 추론, 창의·융합, 의사소통 능력 등을 평가하는 데 활용할 수 있고, 서·논술형 등의 다양한 문항 형태를 활용한다.

- ④ 관찰 평가, 면담 평가, 구술 평가는 학생 개인 및 소집단 관찰 자료, 학생과의 대화 자료, 학생의 발표 자료를 통해 학생의 이해 정도와 사고 방법, 수행 과정 등을 평가하는 방법으로, 의사소통, 태도 및 실천 능력 등을 평가할 때 활용할 수 있다.
- ⑤ 자기 평가는 학생 스스로 자신의 이해와 수행을 평가하는 방법으로, 인공지능에 활용되는 수학 사례와 관련된 문제 해결과 추론 과정의 반성, 자신의 생각 표현, 태도 및 실천 능력 등을 평가할 때 활용할 수 있다.
- ⑥ 동료 평가는 학생들이 동료들을 서로 평가하는 방법으로, 협력 학습 상황에서 학생 개개인의 역할 수행 정도나 집단 활동에 기여한 정도를 평가할 때 활용할 수 있다.

평가 방법 (라) 항은 교수·학습 방법에서 공학적 도구와 교구를 사용하였다면 평가 내용이나 방법에 따라 평가에서도 일관성 있게 이를 이용할 수 있음을 명시하였다.

- (라) 평가 내용이나 방법에 따라 학생에게 계산기, 컴퓨터, 교육용 소프트웨어 등의 공학적 도구와 다양한 교구를 이용할 수 있게 한다.
- ① 공학적 도구로 PC나 모바일 기기를 활용할 수 있으며, 설치형 또는 웹 기반의 소프트웨어를 사용할 수 있다.
 - ② 프로그래밍 능력에 대한 평가보다는 프로그램의 코드에서 수학을 활용하는 능력에 중점을 두어 평가하며, 이를 위해 수학적 요소가 포함되는 프로그래밍 코드를 제공할 수 있다.

III

교수·학습 및 평가 자료 개발 방향

이 교수·학습 및 평가 자료는 2015 개정 수학과 교육과정을 충실히 반영한 것으로서, <인공지능 수학> 수업에서 다양한 교수·학습 방법과 평가 방법을 활용할 수 있도록 돕고자 하였으며 과정 중심 평가가 수업과 연계하여 이루어질 수 있도록 하였다. 또한 2015 개정 수학과 교육과정의 핵심인 수학 교과 역량이 반영된 수업과 평가의 모습을 담아내려 하였다.

01

2015 개정 수학과
교육과정에서
신설된 내용으로
모듈 개발

2015 개정 수학과 교육과정의 <인공지능 수학>은 <수학 I>을 학습한 후, 수학의 지식과 기능을 활용하여 인공지능 속의 수학을 이해하기를 원하는 학생들이 선택할 수 있는 과목이다. <수학 II>의 함수의 극한, 미분계수, <기하>의 벡터, <확률과 통계>의 조건부확률, <고급 수학 I>의 공간 벡터, 행렬과 연관되어 있으며 나아가 인공지능을 이해하고 활용하는 데 있어서 창의적 역량을 갖춘 융합 인재로 성장할 수 있는 기반을 제공한다.

이 자료집에서는 성취기준의 내용을 중심으로 모듈을 선정하였다. 모듈은 동일한 학습 주제로 여러 차시 구성을 묶은 것으로 한 주기의 과정 중심 평가가 이루어질 수 있도록 구성되어 있다. 각 모듈에는 ‘인공지능 발전 과정에서 수학이 어떻게 활용되었을까?’와 같은 제목을 달아 수학 내용뿐 아니라 개발된 교재의 특성을 나타내려 하였다. 이 자료집에 제시된 모듈은 모두 8개이며, 총 15~20차시에 해당되는 분량이다.

핵심 개념	모듈	특징
인공지능과 수학	① 인공지능 발전 과정에서 수학이 어떻게 활용되었을까?	인공지능에 쓰이는 논리와 추론, 전문가 시스템, 진리표, 순서도 작성, 기계학습 경험하기, 분류와 회귀
	② 텍스트 자료는 어떻게 표현할까?	수와 벡터를 이용하여 텍스트 자료를 표현하고, 표, 막대그래프, 단어 구름 등으로 시각화하여 나타내기
자료의 표현	③ 이미지 자료는 어떻게 표현하고 처리할까?	행렬을 이용하여 이미지를 표현하는 방법을 이해하고 행렬의 연산을 활용하여 이미지의 색상 변형하기

분류와 예측	④ 텍스트 자료는 어떻게 분류 할까?	자료 사이의 거리, 코사인 유사도 등의 개념을 바탕으로 문서 분류, 스팸 메일 분류 등 분류 문제 해결하기
	⑤ 이미지 자료는 어떻게 분류 할까?	이미지를 분류하는 방법과 행렬의 곱셈 표현과의 관련성 이해하기
	⑥ 확률을 이용한 예측은 어떻게 할까?	자료를 분석하여 사건이 일어날 확률을 구하고 예측에 이용하기
	⑦ 추세선을 이용한 예측은 어떻게 할까?	자료의 경향성을 추세선으로 나타내고 예측에 이용하기
최적화	⑧ 예측의 정확도를 어떻게 높 일까?	오차를 손실함수로 표현하기, 경사하강법의 원리를 이해하고 경사하강법을 이용하여 오차가 최소인 추세선 찾기

02

수학 교과 역량 반영

수학 교과 역량은 수학 교육을 통해 학생들에게 키워야 하는 능력이지만 다소 추상적일 수 있으므로 교수·학습과 평가에서 구체적인 적용 방안을 보여줄 필요가 있다. 이 자료집에서는 문제 해결, 추론, 창의·융합, 의사소통, 정보 처리, 태도 및 실천의 6가지 수학 교과 역량이 모듈별, 차시별 평가의 채점 기준에 골고루 반영되었다. 특히 교사용 안내서에서 학생들의 활동과 관련한 ‘관련 교과 역량’을 아이콘 과 함께 제시하였다.

03

학생 참여 위주의 수업 실현

2015 개정 수학과 교육과정에 부합하는 교실 수업은 학생들이 능동적으로 참여하고 스스로 수학 개념, 원리, 법칙, 기능을 탐구하는 참여가 활발해야 한다. 이 자료집에서는 학생들이 주도적으로 개념, 원리, 법칙을 유의미하게 학습하고 논리적 사고를 충분히 경험할 수 있는 활동을 많이 포함하도록 하였다. 협력 학습이나 탐구 학습 등 다양한 교수·학습 방법을 적용하여 깊이 있는 배움이 가능한 수업을 실현하려 하였다. 이에 2015 개정 수학과 교육과정의 ‘교수·학습 및 평가의 방향’에서 다양한 교수·학습 방법을 모듈에 되도록 포함하려 하였다.

04

과정 중심 평가 반영

교수·학습 방법	적용 방법
설명식 교수	교사가 수업 중 설명과 시연을 함.
탐구 학습	학생이 수학 개념, 원리, 법칙을 발견하고 구성하게 함.
프로젝트 학습	모둠을 구성하여 특정 주제에 대해 계획을 수립하고 수행하여 결과물을 산출하고 발표하게 함.
토의·토론 학습	특정 주제에 대해 협의하거나 논의하게 함.
협력 학습	모둠 활동을 통해 공동의 학습 목표에 도달하게 함.
매체 및 도구 활용 학습	매체와 도구를 활용하여 학습의 효율성과 다양성을 도모함.

2015 개정 수학과 교육과정은 학습 결과 평가뿐 아니라 과정 중심 평가를 실시할 것을 권고하고 있다. 이 자료는 수업을 위한 자료에서 그치는 것이 아니라 수업과 연동되어 진행되는 과정 중심 평가가 어떤 모습일지를 담아내려 했다. 이를 위해 다음과 같은 내용을 자료에 포함하였다.

- 모듈별 평가 계획
- 차시별 평가 결과의 피드백
- 교사용 안내서 위쪽에 ‘포트폴리오 평가 반영’과 같은 표기
- 평가 결과에서 평가 방법에 따른 성취 수준별 특징, 학생 개별 학교 생활 기록부 기재

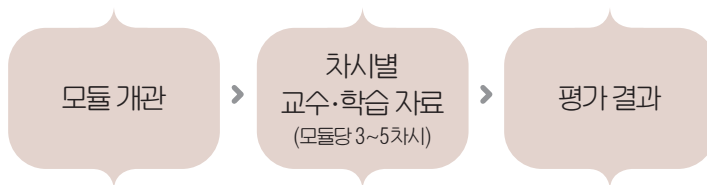
그리고 과정 중심 평가에서 활용할 수 있는 다양한 평가 방법을 모듈에 적용하였다. 인지적 영역뿐 아니라 정의적 영역인 태도 및 실천 역량까지도 평가할 수 있는 방안을 제시하여, 학습 과정과 결과, 지식과 역량을 균형 있게 평가하는 예시를 제안하려 하였다.

평가 방법	적용 방법
지필 평가	수업 전후 또는 중간에 선택형, 단답형, 서·논술형 문항 형태로 평가함.
프로젝트 평가	특정 주제에 대해 계획을 수립하고 수행하여 결과물을 산출하는 과정과 그 결과를 발표하는 내용을 평가함.
포트폴리오 평가	일정 기간 수행한 수학 학습 과정과 결과를 평가함.
관찰, 면담, 구술 평가	학생 또는 모둠을 관찰, 학생과의 대화, 학생 발표를 평가함.
자기 평가	학생 스스로 자신의 이해와 수행을 평가함(위의 평가 방법과 중복될 수 있음).
동료 평가	모둠 내 동료의 수행을 평가함(위의 평가 방법과 중복될 수 있음).

05

모듈의 구성

이 자료집은 고등학교 <인공지능 수학>의 4개 핵심 개념에 대해 총 8개의 모듈을 제시하였다. 모듈은 한 주기의 과정 중심 평가가 이루어질 수 있는 주제로 구성되어 있다. 각 모듈은 모듈 개관, 차시별 교수·학습 자료(교사용 안내서), 평가 결과로 이루어진다.



모듈의 구성

모듈 개관

- 모듈 자료 개발 방향에 대한 소개
- 모듈과 관련된 교육과정 성취기준, 모듈 설계 의도, 교과 역량, 학습 요소, 교수·학습 방법, 교수·학습 유의 사항, 평가 방법, 평가 방법 및 유의 사항, 차시별 주제 및 교수·학습 활동과 평가 활동 제시

차시별 교수·학습 자료

- 차시 소개(학습 목표, 수업의 의도, 준비물) 및 수업 지도안(수업 단계별 교수·학습 활동과 평가 활동)
- 교사용 안내서



교사용

평가 결과

- 평가 방법별 학생들의 성취 수준
- 학교 생활 기록부 기재 예시

자료집 활용 방안

이 자료집의 학생용 활동지는 출력하거나 복사해서 쓸 수 있도록 개발되었다. 활동지가 교과서를 대신하는 자료는 아니므로 때에 따라 활동지와 교과서를 함께 활용해야 하는 경우가 있음에 유의해야 한다.

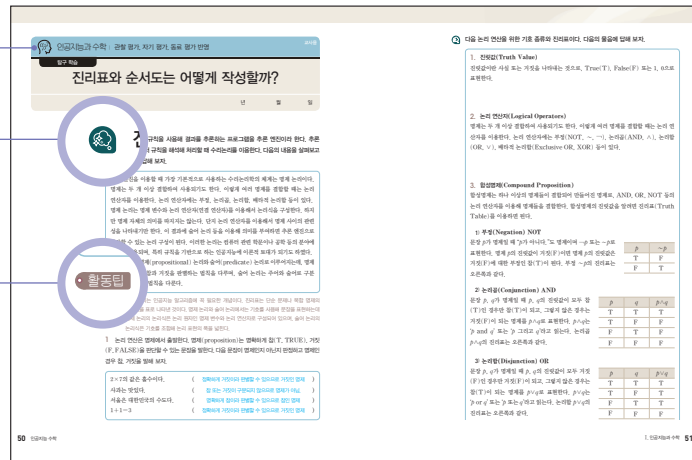
또한 학생용 활동지로 활용할 때 이 자료집에 제시된 교사용 안내서의 예시 답과 활동 팁 등을 삭제한 후 필요한 부분만 남겨서 활용할 수 있다.

자료집 활용 시 유의 사항

한 가지의 평가 방법에 대해서도 평가 도구는 다양할 수 있다. 예를 들어, 학생들의 개별 관찰 평가를 실시할 때 관찰 기록의 방법이나 형식이 수업 내용에 따라 다를 수 있다. 따라서 이 자료집에서는 통일성 있는 양식을 만드는 대신 다양한 평가 도구의 활용 모습을 보여줄 수 있도록 하였다.

아이콘 설명

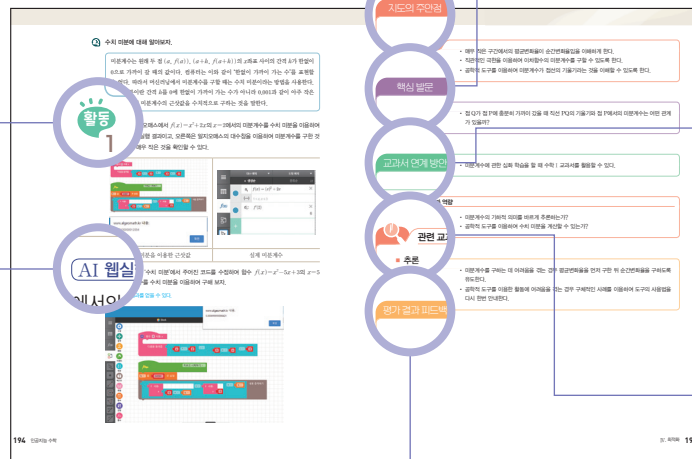
- 교사용 안내서
- 주제 도입 및 내용 설명
- 활동팁
- 수업 중 교사가 알아두면 좋은 내용



지도의 주안점

지도 시 교사가 강조할 점 또는 수업 방법에 대한 안내

- 활동
- AI 웹실습실 또는 공학적 도구 활용
- AI 웹실습실
- 인공지능 수학 모듈형
- 교재 학습 내용에 대하여 학생들이 직접 체험할 수 있는 웹 공간



핵심 발문

차시 수업에서 학생들에게 강조해서 지도하고 싶은 내용을 제시한 발문

교과서 연계 방안

교수·학습 자료와 교과서를 함께 사용할 때 각각의 활용 시점에 대한 안내 또는 학생들의 이해를 제고 하기 위해 교과서에서 추가적으로 다루어야 하는 내용

관련 교과 역량

평가 결과의 피드백

차시 학습 후 학생들의 학습을 증진하고 다음 수업을 위해 준비해야 하는 사항

AI 웹실습실 <http://askmath.re.kr>

'이미지 처리', '텍스트 분석', '예측과 최적화'로 구성되어 있다.

이미지 처리에서는 이미지 자료의 변형에 따른 행렬의 연산이 변하는 활동을 체험하고 이해할 수 있도록 구성하였다.

텍스트 분석에서는 실제 접하는 텍스트 자료를 수와 벡터로 표현하고 계산해봄으로써 텍스트 자료의 표현 방법을 이해할 수 있도록 구성하였다.

예측과 최적화에서는 알지오매스를 활용하여 틱택토, 산점도와 추세선, 미분계수의 기하적 의미, 수치 미분, 경사하강법을 이해할 수 있도록 구성하였다.

I 인공지능과 수학

모듈① 인공지능 발전 과정에서 수학이 어떻게 활용되었을까?

II 자료의 표현

모듈② 텍스트 자료는 어떻게 표현할까?

모듈③ 이미지 자료는 어떻게 표현하고 처리할까?

III 분류와 예측

모듈④ 텍스트 자료는 어떻게 분류할까?

모듈⑤ 이미지 자료는 어떻게 분류할까?

모듈⑥ 확률을 이용한 예측은 어떻게 할까?

모듈⑦ 추세를 이용한 예측은 어떻게 할까?

IV 최적화

모듈⑧ 예측의 정확도를 어떻게 높일까?

2015 개정 교육과정 교수·학습 자료
인공지능 수학 **교사용**

제 2 부

교수·학습 및 평가 자료

I

인공지능과 수학

1. 영역의 개관

인공지능에서 수학이 활용되는 다양한 사례를 경험함으로써, 인공지능과 수학의 연관성을 이해하고 수학의 유용성과 가치를 느낄 수 있다.

가 2015 개정 수학과 교육과정 내용

1. 성취기준 및 학습 요소

인공지능과 수학 영역에 제시된 성취기준 및 학습 요소는 다음과 같다.

□ 인공지능과 관련된 수학

[12인수01-01] 인공지능의 발전에 기여한 역사적 사례에서 수학이 어떻게 활용되었는지를 이해한다.

[12인수01-02] 인공지능에 수학이 활용되는 다양한 예를 찾을 수 있다.

관련 학습 요소

• 진리표, 순서도

각각의 학습 요소를 간단히 설명하면 다음과 같다.

진리표

진리표는 단순문제나 복합명제의 진릿값을 표로 나타낸 것이다. 명제는 ‘참이나 거짓으로 구분할 수 있는 문장이나 수식’을 뜻하는 것으로 수학적 논리(Mathematical Logic)는 명제가 참인지 거짓인지 판별한다. 인공지능은 프로그램을 수행하여 올바른 결과를 얻으려면 수학적 논리를 활용하는데, 명제를 합성하는 데 사용하는 다양한 논리 연산자(Logical Operator)에 의한 진릿값을 활용하여 다양한 추론을 한다.

순서도

순서도는 알고리즘을 알기 쉽게 기호와 그림으로 나타낸 것이며, 알고리즘이란 어떤 문제를 해결하기 위한 유한 번의 절차를 말한다. 즉, 순서도란 처리하고자 하는 문제를 분석하여 그 처리 순서를 단계화시켜 상호 간의 관계를 일정한 기호를 사용하여 일목요연하게 나타낸 그림으로 알기 쉽게 도식화하는 것을 의미한다. 순서도는 프로그래밍을 할 때 많이 활용되고 있다.

2. 교수·학습 방법 및 유의사항

인공지능과 수학 영역의 교수·학습 방법 및 유의사항은 다음과 같다.

- 수학이 인공지능의 발전에 기여한 역사적 사례로 논리, 진리표, 알고리즘, 순서도, XOR문제, 전문가 시스템, 딥러닝 등을 소개할 수 있다.

‘1950년대 프랭크 로젠블랫, 퍼셉트론, 논리, 진리표, XOR문제’, ‘1960년대 전문가 시스템’, ‘1980년대 인공신경망’, ‘1990년대 수학 개념 본격 도입(선형대수학, 확률과 통계, 최적화 이론, 이산수학 등)’, ‘2000년대 이후 딥러닝’을 이해한다.

- 실생활에서 활용되는 인공지능에서 수학이 활용되는 다양한 사례를 찾아보는 활동을 하게 한다.

생활 속 인공지능의 실생활 사례 예시로 동영상 추천 시스템, 안면 인식 시스템, 댓글 분석(영화감상평 등), 인공지능 스피커와 주파수 그래프 등이 있다.

3. 평가 방법 및 유의 사항

인공지능과 수학 영역의 평가 방법 및 유의 사항은 다음과 같다.

- 인공지능과 관련된 수학에 대한 평가를 할 때는 학생들의 흥미, 자신감, 수학의 가치 인식, 학습 태도 등 정의적 영역의 평가를 할 수 있다.

인공지능(AI), 데이터 사이언스 등 첨단기술의 주요 기저로 수학이 활용됨을 이해하고 인공지능에 수학이 활용되는 예들을 접함으로써 수학에 대한 가치와 유용성을 알 수 있다. <인공지능 수학> 과목의 첫 번째 핵심 개념인 ‘인공지능과 수학’에서는 이후 학습하게 될 자료의 표현이나 분류와 예측, 최적화에서 다룰 내용의 개요와 필요성을 이해하고 동기를 유발하도록 정의적 영역의 평가를 한다.

2. 모듈①

인공지능 발전 과정에서 수학이 어떻게 활용되었을까?

가 모듈 개관

성취기준	12인수01-01 인공지능의 발전에 기여한 역사적 사례에서 수학이 어떻게 활용되었는지를 이해한다. 12인수01-02 인공지능에 수학이 활용되는 다양한 예를 찾을 수 있다.
모듈 설계 의도	성취기준의 의미 인공지능이 발전한 역사적 사례를 바탕으로 인공지능과 밀접하게 연계되어 있는 수학 개념의 예를 파악한다. 수학적 활동을 통해 인공지능의 원리를 탐구함으로써 수학의 필요성과 유용성을 이해한다. 학습 활동 계획 이 모듈에서는 ‘인공지능 발전 과정에서 수학이 어떻게 활용되었을까?’라는 주제로 인공지능에 쓰이는 논리와 추론, 전문가 시스템을 다루며, 진리표, 순서도를 작성하고, 기계학습 경험하기를 통해 인공지능에 활용되는 간단한 수학을 탐구하고 찾아볼 수 있도록 하면서 생활 속 인공지능에서 활용되는 수학을 학습자 중심으로 탐구할 수 있게 활동이 구성된다. 기대하는 학습 결과 모듈별 탐구 학습을 통해 인공지능의 개념을 습득하게 될 것이다. 더 나아가 인공지능에서의 수학의 활용과 중요성을 이해하면서 수학에 대한 가치를 알고 흥미를 가질 수 있을 것이다.
교과 역량	의사소통, 정보 처리, 태도 및 실천
학습 요소	진리표, 순서도
교수 · 학습 방법	탐구 학습
교수 · 학습 방법 및 유의사항	• 수학이 인공지능의 발전에 기여한 역사적 사례로 논리, 진리표, 알고리즘, 순서도, XOR 문제, 전문가 시스템, 딥러닝 등을 소개할 수 있다. • 인공지능과 관련된 수학에 대한 평가를 할 때는 학생들의 흥미, 자신감, 수학의 가치 인식, 학습 태도 등 정의적 영역의 평가를 할 수 있다.
평가 방법	관찰 평가, 자기 평가 및 동료 평가
평가 방법 및 유의 사항	• 인공지능에서 활용되는 논리표와 순서도를 올바르게 이해하고, 표현할 수 있는지 평가한다.

차시	주제	교수·학습 활동	평가 활동
1	인공지능의 역사, 논리, 전문가 시스템	탐구 학습 논리와 추론 - 인공지능에 쓰이는 논리와 추론 알아보기 - 전문가 시스템 알아보기	관찰 평가 - 논리와 추론의 의미를 이해하고 새로운 문제 상황에 적용할 수 있는가? 자기 평가 및 동료 평가 - 전문가 시스템에 대한 흥미는 어떠한가? - 모둠 활동 시 태도는 어떠한가?
2	진리표, 순서도에 대한 내용 및 활동	탐구 학습 진리표와 순서도 - 진리표 작성하기 - 순서도 작성하기	관찰 학습 - 진리표와 순서도의 의미를 이해하고 문제를 해결할 수 있는가? 자기 평가 및 동료 평가 - 진리표와 순서도에 대한 흥미는 어떠한가? - 모둠 활동 시 태도는 어떠한가?
3	머신러닝에 대한 개관 소개, 실생활 인공지능에서의 수학	탐구 학습 분류와 회귀 - 생활 속 인공지능 조사하기 - 기계학습 경험하기 - 인공지능에 활용되는 수학적 개념 알아보기	관찰 학습 - 생활 속 인공지능을 조사하고 발표하였는가? - 인공지능에서 분류와 회귀를 수행하는 자료의 특징을 이해하고 적절한 사례를 제시할 수 있는가? 자기 평가 및 동료 평가 - 인공지능에 대한 흥미는 어떠한가? - 모둠 활동 시 태도는 어떠한가?

1차시 | 인공지능 역사와 수학

학습 목표	• 인공지능의 역사에서 다양한 수학이 활용되었음을 탐구할 수 있다. • 논리와 전문가 시스템을 이해할 수 있다.
수업의 의도	수업 의도 대중매체에서 소개되고 있는 인공지능 사례를 통해 인공지능의 개념 및 실생활에서의 활용에 대하여 이해한다. 또한 수학이 인공지능에 활용되고 있음을 알고 수학 가치와 유용성을 인식할 수 있게 한다. 수업의 주안점 다양한 방법을 통해 인공지능에 대하여 이해하고, 인공지능이 발전하는 과정에서 수학의 원리가 활용되고 있음을 알고 의사소통할 수 있게 한다. 수업 활동 시 유의점 다양한 인공지능 개념을 학생의 수준에 맞추어 이해할 수 있도록 하며, 탐구 활동 위주의 수업을 통해 지식 위주의 개념 전달이 되지 않도록 교사는 적절한 발문을 제공한다. 평가의 주안점 다양한 추론 절차를 올바르게 이해한 후에 올바른 추론을 해낼 수 있는지와 자신의 결과를 오류 없이 표현할 수 있는지 평가한다.
준비물	활동지, 계산기, 자기 평가 및 동료 평가지

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	• 2~4인 1조로 모둠 구성 • 활동지, 자기 평가 및 동료 평가지 배부 • 인공지능 관련 기사 소개	인공지능에 대한 흥미 등에 대한 정의적 평가
수행	활동	• 인공지능 역사 찾아보기 • 추론의 종류에 관해 토론하기 • 추론 표 완성하기 • ‘상자 썩기’에 대한 규칙 탐구하기 • ‘물통 문제’에 대한 규칙 탐구하기 • ‘선교사와 식인종 문제’에 대한 규칙 탐구하기	관찰 평가
정리		• 자기 평가 및 동료 평가지 작성	자기 평가, 동료 평가



인공지능의 역사와 수학

년

월

일



다음 글을 읽고 모둠별로 논의해 보자.

4차 산업혁명 시대, 인공지능의 역할은?

최근 ‘다 아는 단어인데도 뜻이 확실히 뭔지 애매한’ 말들이 범람하고 있다. 대표적인 게 4차 산업혁명이나 인공지능 같은 것들이다. 산업혁명이라고 하면 누구나 증기기관이나 전기에너지에 의한 대량 생산을 떠올릴 것이다. 흔히 4차 산업혁명의 핵심 기술은 인공지능이라고들 한다.

Artificial intelligence

지능은 ‘외부를 인식하고 추론하며 적응하는 능력’으로 정의된다. 하지만 인간조차 그런 기능이 어떻게 발현되는지 모르는 상태에서 지능을 만드는 건 결코 쉽지 않다.

하지만 인간조차 그런 기능이 어떻게 발현되는지 모르는 상태에서 전통적 방법으로 지능을 만드는 건 결코 쉽지 않다. 일반적으로 인간에게 “어렵다”고 여겨지는 행위, 이를테면 큰 수를 곱하거나 미적분 하기, 체스나 바둑 두기, 금융 시장에서 투자 결정 내리기 등은 컴퓨터로 비교적 쉽게 처리할 수 있다. 반면, 사진 보고 개인지 고양이인지 구별하기, 동화책 읽고 그 내용 이해하기 등 인간이 쉽게 해내는 행위를 컴퓨터로 구현해내긴 어렵다. 오늘날 관련 기술 발달로 예전에 비해 정확도가 매우 높으면서 가격이 대폭 저렴해진 센서를 다양한 지점에 부착할 수 있게 됐다. 또한 정보통신 기술의 비약적 발전에 힘입어 자동 생산과 지능형 시스템 구축을 위한 요소 기술도 완비됐다. 사물인터넷(IoT)은 이 같은 센서와 네트워크를 활용, 각종 빅데이터를 쏟아낸다. 결국 이를 효과적으로 관리, 분석하는 소프트웨어가 필요한데 그게 바로 인공지능이다.

인공지능 하면 대부분 막연하게 공상 과학 영화를 떠올린다. 하지만 현실에서 인공지능 소프트웨어는 제조 현장에서 쏟아지는 빅데이터를 지능적으로 처리, 제조업 분야의 효율성과 서비스업 분야의 편의성을 높이는 데 활용된다. 인공지능은 전문가 시스템과 신경망, 딥러닝(deep learning)에 이르기까지 실로 다양한 방법을 시도하고 있다. 혹자는 인공지능을 “어떤 문제도 해결할 수 있는 마법의 열쇠”로 여긴다. 하지만 인공지능으로 성공을 거두려면 풀고자 하는 문제를 잘 이해하고 정형화하는 노력이 선행돼야 한다.

4차 산업혁명에는 사람과 사물, 공간을 초(超)연결해 산업 구조와 사회 체계에 혁신을 일으키는 게 골자다. 인공지능을 ‘막연한 미래 기술’이 아니라 ‘4차 산업혁명 시대를 선도하는 실질적 기술’로 활용하려면 어떻게 해야 할까? 인공지능 분야엔 수십 가지의 다양한 기술이 존재한다. 각 방법에 따른 특징점이 존재하므로 실제로 이들을 복합적으로 활용하면 시너지 효과를 거둘 수 있다. 이렇게 볼 때 인공지능 기술 자체도 중요하지만 그것 못지않게 풀고자 하는 문제의 특성을 체계적으로 분석, 정리하고 각 부분에 적합한 인공지능 기술을 복합적으로 활용하는 자세가 필요하다.

출처: 조성배, 4차 산업혁명 시대, 인공지능의 역할은?, 2018.
(<https://news.samsung.com/kr>)

1 인공지능이 무엇인지 말해 보자.

• **활동팁** 1970년에 수학자 마빈 민스키는 인간의 뇌처럼 정보를 저장했다가 꺼내 쓸 수 있는 기계인 컴퓨터 에니악을 제시하면서 컴퓨터의 전자부품을 뇌세포처럼 연결하고 적절한 프로그램을 넣어 주면 뇌와 비슷한 활동을 하리라 생각하였고 인간이 만들어낸 지능이란 의미로 인공지능이란 아이디어를 제시하였다.

■ 지능이란?

- 본능적이나 자동적으로 행동하는 대신에, 생각하고 이해하여 행동하는 능력이며, 인간이 사물을 이해하고 학습(learning)하는 능력과 어떤 문제가 주어졌을 때, 합리적으로 사고하여 문제를 해결하는 능력(problem solving)과 연관되어 있다.

■ 인공지능(AI, Artificial Intelligence)이란?

- 인공적으로 만든 지능으로 “인간의 인지적인 기능을 흉내 내서 문제를 해결하기 위해 학습하고 이해하는 기계(컴퓨터)”라 할 수 있으며, 인간이 가지고 있는 지적 능력을 인공적으로 구현하는 기술로, 인간의 지능으로 할 수 있는 사고·학습·자기 계발 등을 컴퓨터로 구현하는 방법에 대해 연구하는 ICT 기술의 한 분야이다.
 - 인간처럼 사고하기(Thinking Humanly) - 인지 과학, 신경망
 - 합리적으로 사고하기(Thinking Rationally) - 논리학, 추론
 - 인간처럼 행동하기(Acting Humanly) - 튜링 테스트, 로봇 공학
 - 합리적으로 행동하기(Acting Rationally) - 에이전트: 목표를 성취하기 위한 행동, 추론 포함
- 강한 인공지능(Strong AI)
 - 사람과 같은 수준의 지능(학습해서 스스로 진화하는 인공지능)
 - 마음을 가지고 사람처럼 느끼면서 지능적으로 행동하는 기계
 - 추론, 문제 해결, 판단, 계획, 의사소통, 자의의식(self-awareness), 감정(sentiment), 지혜(sapience), 양심(conscience)의 기능을 갖는다.
 - 키스멧(Kismet): 감정 표현 얼굴 로봇
- 약한 인공지능(Weak AI, narrow AI)
 - 특정 문제를 해결하는 지능적 행동
 - 사람의 지능적 행동을 흉내 낼 수 있는 수준(인간이 할 수 있는 일 중 일부만 대신한다.)
 - 대부분의 인공지능 접근 방향

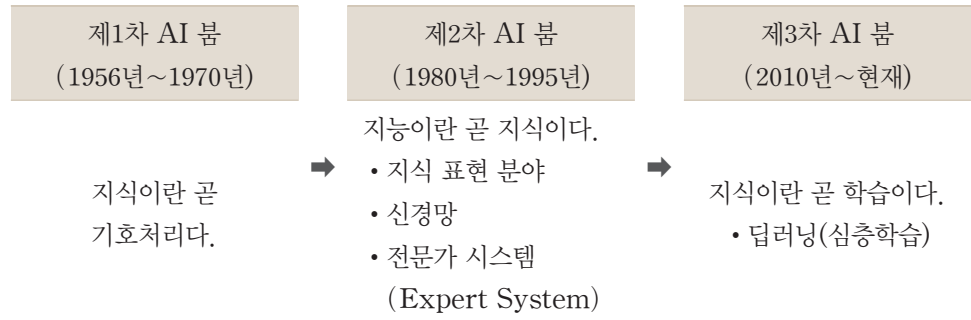
2 인공지능이 우리 생활에 어떤 영향을 끼칠 것인지 논의해 보자.

- 최근 들어 인공지능 기술은 예술이나 문학 분야 등에도 도전하며 다방면에 걸쳐 인간 생활에 활용되고 있다.
 - 유튜브(YouTube)에서 인공지능이 동영상을 추천하고, 인터넷으로 신문 기사를 읽는 도중에 중간 광고가 나타난다(여러 번 검색하며 관심을 가졌던 주제와 관련된 광고). 그 외 도서 검색 경험을 바탕으로 한 추천 광고 등으로 개인에게 맞춤형 소비가 이루어지도록 하고 있다.
 - 인공지능 기술로 얼굴 인식과 사물 인식을 실용화하고 있으며 이를 통해 신분증 인증 시스템이 간편화되고 있다.
 - 스포츠 경기에서 심판의 공정성 시비를 인공지능 ‘로봇 심판’이 판정함으로써 시비나 불만이 줄어들고 있다.
 - 인공지능 스피커로 집안의 기기를 음성으로 제어하고 통신망에 연결된 가전기기를 제어할 수 있다.
 - 법조인의 업무를 보조하고 판례 정보를 제공하고, ‘왓슨’과 같은 인공지능이 암 발견과 최적의 암 환자 치료를 제안하고 있는 등 법률이나 의학에 인공지능이 쓰이면서 인간이 할 일을 더 정확히 수행하면서 우리 생활의 편리성과 효율성을 제공하고 있다.
 - 기사 작성, 시와 소설을 쓰고, 작곡하고, 그림을 그리는 단계에 진입하면서 머지않은 미래에는 예술 장르에 본격적으로 도전 가능하며 인공지능은 언젠가는 창의성(creativity)에도 도전할 것이라고 예측되고 있다.
- 인공지능이 발달하면 그로 인한 여러 가지 부작용 역시 예상된다.
 - 인공지능의 발달은 인간의 일자리를 잃게 할 수 있다.
 - 인공지능이 발달한 사회에서 윤리적인 문제가 발생할 수 있다.
 - 인공지능에 의지하다 보면 인간의 판단이 흐려질 수도 있다.
 - 인공지능에 의지하는 습관에 종속될 가능성도 없지 않다.
 - 인공지능에 의해서 감시를 받을 가능성이 크다.



오늘 배울 인공지능의 역사에 관하여 알아보자.

1 인터넷을 참고하여 인공지능이 발전해 나간 시기별 특징에 대해 조사해 보자.



연도	구분	내용
1943년~1956년	준비기	• 명제 논리의 개념 • 튜링 테스트 • 최초의 체스 프로그램 작성
1956년~1974년	요람기 (제1차 AI 붐)	• ‘인공지능’이란 이름 탄생 • ‘마크 1 퍼셉트론’ 신경망 개발 • Lisp 언어 개발 • ELIZA 발표 • ‘퍼셉트론(Perceptrons)’ 책 출간
1974년~1980년	첫 번째 겨울기	• ‘퍼셉트론’ 이후 신경망 쇠퇴로 연구 자금 급격히 줄어듦.
1980년~1987년	발전기 (제2차 AI 붐)	• ‘XCON’이란 전문가 시스템 개발 • 제5세대 컴퓨터 프로젝트 수행 • 다층 신경망으로 부활
1987년~1993년	두 번째 겨울기	• 제한적 성능과 컴퓨팅 파워 부족 • 신경망의 정체 • 300개 이상 인공지능 시장 붕괴 • 인공지능 관련 정부 연구 자금 고갈
1993년~2011년	안정기	• 세계 체스 챔피언 • 제퍼디(Jeopardy!) 퀴즈쇼 챔피언
2011년~현재	부흥기 (제3차 AI 붐)	• 딥러닝 기반의 ImageNet • 고양이 인식 구글 프로젝트 • 딥러닝 기법의 세계 바둑 챔피언

 인공지능에 쓰이는 논리와 추론에 대해 살펴보자.

추론이란 ‘알려진 사실이나 명제를 근거로 삼아 미지의 사실에 관한 판단이나 결론을 이끌어 내는 사고 과정’이라 할 수 있다. 어떤 명제가 참인 것을 근거로 다른 명제가 참임을 유도하는 이러한 방식은 우리가 어떤 글의 앞뒤 사실로 미루어 추론하는 것과 같은 원리이다. 인공지능은 주어진 몇 가지 사실들과 추론의 규칙들을 가지고 새로운 사실을 찾아내어 적용하는 방식을 활용하고 있다. 논리형 프로그래밍인 프로로그(Prolog)는 추론에 주로 쓰이는 대표적인 프로그래밍 언어이기도 하다.



1 다음 추론의 종류와 방식을 확인하고 다음의 물음에 답해 보자.

추론의 종류	방식
귀납추론 (Induction)	개별적인 사실들로부터 일반적인 결론을 이끌어 내는 추론 방식 - 연역 추론과 같은 논리적 필연성이 적다. - 새로운 지식이나 이론의 발견과 확장에 쓰이는 추론 방법
연역추론 (Deduction)	전제로부터 다른 결론을 도출해내는 추론 방식 - 가장 널리 쓰이는 추론 방법 - 최초의 대전제가 결론을 이끌어내는 가장 중요한 근거 - 확장성은 부족하나 논리의 일관성이 장점 - 연역 추론 중 삼단 논법이 잘 알려졌다.
유비추론 (Analogy)	A라는 대상과 B라는 대상이 서로 유사할 때, A에서 성립하는 성질과 유사한 성질이 B에서도 성립할 것이라고 주장하는 방식 - 사다리꼴 넓이 → 각뿔대의 부피 - 유추를 관계적 성질에 주목하므로 패턴이나 법칙을 다루는 학습에 강력한 사고 도구가 된다.

1) 다음을 추론하여 표를 채워 보자.

알려진 사실		일반적 결론
사실 #1	까마귀 1은 까맣다.	모든 까마귀는 까맣다.
사실 #2	까마귀 2는 까맣다.	
사실 #3	까마귀 3은 까맣다.	
사실 #4	까마귀 4는 까맣다.	
⋮	⋮	
사실 #N	까마귀 N은 까맣다.	

2) 삼단 논법에 따라 결론을 추론해 보자.

지식	$A \rightarrow B$	소크라테스는 사람이다.
사실	$B \rightarrow C$	사람은 모두 죽는다
결론	$C \rightarrow A$	소크라테스는 죽는다

3) 다음의 주어진 추론 전략을 가지고 결론을 추론해 보자.

지식	$A \rightarrow B$	오늘이 월요일이면, 홍길동은 일하러 간다.
사실	A	오늘은 월요일이다.
결론	B	홍길동은 일하러 간다.

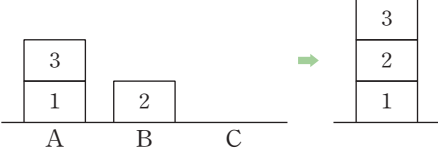
지식	$A \rightarrow B$	오늘이 월요일이면, 홍길동은 일하러 간다.
사실	$\sim B$	홍길동은 일하러 가지 않았다.
결론	$\sim A$	오늘은 월요일이 아니다.

2 인공지능이 문제를 해결하는 단계에 관하여 확인하고 다음의 물음에 답해 보자.

■ 인공지능 문제 해결을 위한 주요 단계

- 문제를 명확하게 정의
- 문제에 대한 철저한 분석
- 정해진 제약 조건이나 규칙이 있는 경우 규칙의 적용에 대한 검증
- 최적의 기법 선택
- 결과가 나오면 과정에 문제가 없는지 분석 검토

1) 초기의 인공지능 적용 문제인 '상자 쌓기'에 대한 규칙이다. 상자의 이동 횟수와 방법을 구해 보자.

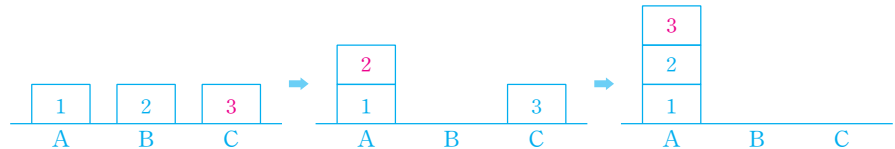
목표 상태	규칙
	• 왼쪽에 상자를 하나씩 옮겨 목표 상태로 만든다. • 작은 번호 상자가 큰 번호 상자 밑으로 가도록 한다. • 상자는 다른 상자 위나 바닥 C에 놓을 수 있다.

다음 절차대로 가능하며 다음과 같이 그림으로 표현할 수 있다.

- 먼저 3번 상자를 비어 있는 공간 C에 옮긴다.

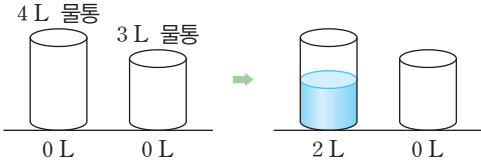


- A의 1번 상자 위에 2번, 3번 상자를 차례로 쌓는다.

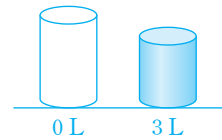


단 3회의 이동으로 목표 상태를 만들 수 있다.

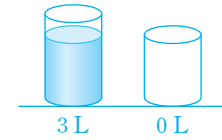
2) 다음은 인공지능과 관련된 ‘물통 문제’에 대한 규칙이다. 4 L 짜리 물통에다 2 L의 물을 채우는 방법을 구해 보자.

목표 상태	규칙
4 L와 3 L 물통이 각각 하나씩 있다. (단, 물통에는 용량을 나타내는 어떠한 표시도 없다.)  4 L 물통 0 L 3 L 물통 0 L → 4 L 물통 2 L 3 L 물통 0 L	• 물통에 물을 가득 채우기 • 물통의 물을 전부 다 비우기 • 물통에 남은 물을 다른 물통에다 붓기 • 물통의 물을 다른 물통이 가득 찰 때까지 붓기

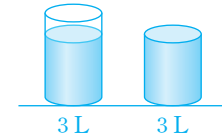
❶ 3 L 물통에 물을 가득 채운다.



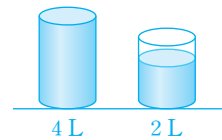
❷ 3 L 물통에 가득 담긴 물을 4 L 물통에 모두 붓는다.



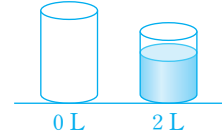
❸ 비어 있는 3 L 물통에다 다시 물을 가득 채운다.



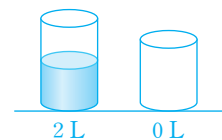
❹ 3 L 물통에 가득 담긴 물을 4 L 물통이 가득 찰 때까지 붓는다.



❺ 4 L 물통의 물을 모두 비운다.



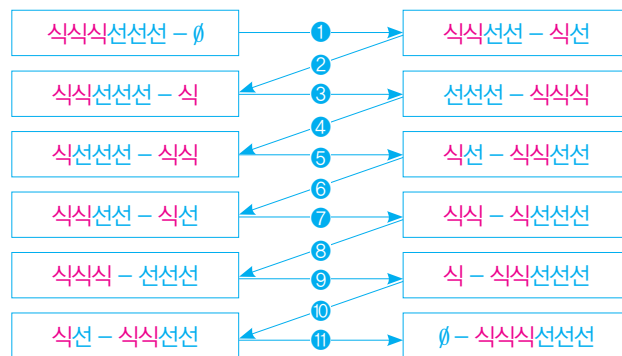
❻ 3 L 물통에 담긴 물을 4 L 물통에 모두 붓는다.



3) 다음은 인공지능 규칙 적용을 이용하기 위한 '선교사와 식인종 문제'에 대한 규칙이다. 모두 건널 수 있는 방법을 구해 보자.

목표 상태	규칙
3명의 선교사, 3명의 식인종과 1척의 배가 강 왼쪽에 있다.  - 선교사와 식인종 모두 배를 타고 강을 건너야 한다.	- 배에는 최대 2명까지 탈 수 있다. - 선교사와 식인종 누구나 배를 조종할 수 있다. - 식인종 수가 선교사 수보다 많을 경우 식인종이 선교사를 해친다.

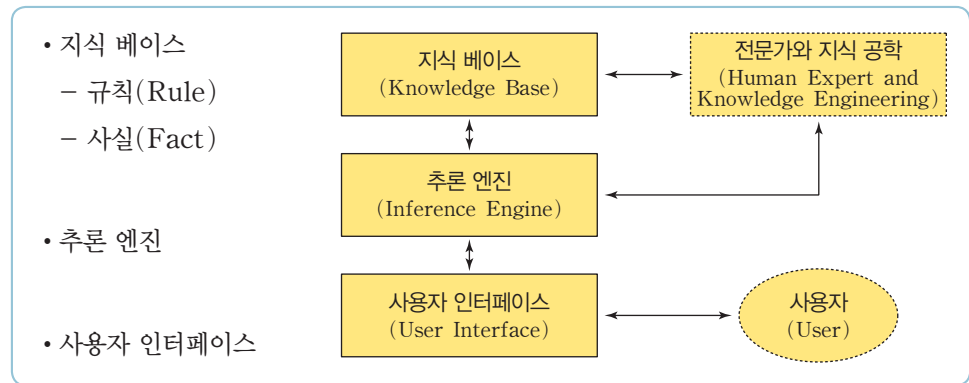
- ① 선교사 1명과 식인종 1명이 왼쪽에서 오른쪽으로 이동한다.
- ② 선교사 1명이 오른쪽에서 왼쪽으로 이동한다.
- ③ 식인종 2명이 왼쪽에서 오른쪽으로 이동한다.
- ④ 식인종 1명이 오른쪽에서 왼쪽으로 이동한다.
- ⑤ 선교사 2명이 왼쪽에서 오른쪽으로 이동한다.
- ⑥ 선교사 1명과 식인종 1명이 오른쪽에서 왼쪽으로 이동한다.
- ⑦ 선교사 2명이 왼쪽에서 오른쪽으로 이동한다.
- ⑧ 식인종 1명이 오른쪽에서 왼쪽으로 이동한다.
- ⑨ 식인종 2명이 왼쪽에서 오른쪽으로 이동한다.
- ⑩ 선교사 1명이 오른쪽에서 왼쪽으로 이동한다.
- ⑪ 선교사 1명과 식인종 1명이 왼쪽에서 오른쪽으로 이동한다.





인공지능에 쓰이는 전문가 시스템(Expert System)의 구성에 대해 알아보자.

1 인터넷을 참고하여 전문가 시스템의 특징에 대해 조사하고, 전문가 시스템의 장점과 단점이 무엇인지 생각해 보자.



■ 전문가 시스템의 특징

- 인공지능 기술의 응용분야 중에서 가장 활발하게 응용되고 있는 분야이다.
- 인간이 특정분야에 대하여 가지고 있는 전문적인 지식을 정리하고 표현하여 컴퓨터에 기억시킴으로써, 일반인도 이 전문지식을 이용할 수 있도록 하는 시스템이다.
- 지식 기반 시스템(Knowledge-based System)으로 특수한 영역에서 인간 전문가의 의사 결정 능력을 흉내 내는 컴퓨터 시스템을 의미한다.
- 특정 제한된 문제 영역에 대해 전문가 수준의 해법을 제공하는 것이다.
- 인간의 사고체계를 연구한 추론 기법들도 중요하지만, 실제적인 문제를 해결하기 위해서는 일반적 해보다 제한된 도메인의 문제에 역량을 집중하는 것이 필요하다.
- 문제보다 제한된 도메인의 문제에 역량을 집중하는 것이 필요하다.
- 특정한 추론 기법보다 보유했던 '지식'이 더 중요하다.
- 유전학자 레더버그(Joshua Lederberg)의 외계 생명체 연구를 돕기 위해 화학 분야에서 개발한 것이다.
- 의료 진단 시스템, 설계 시스템 등이 있다.

■ 전문가 시스템의 장점과 단점

장점	단점
• IF-THEN의 규칙 기반 지식 표현은 인간이 이해하기에 자연스러운 방법이다.	• 정해진 지식 내에서만 동작하며, 시스템이 동작하는 동안 쌓인 경험이 학습되지는 않는다.
• 지식은 항상 IF-THEN 구조로 표현되기에 규칙은 독립적으로 표현 가능하다.	• 규칙이 많아지면 실행 속도가 느려지며, 탐색이 비효율적이게 된다.
• 문제에 의존성이 있는 지식 베이스와는 달리 추론 엔진은 독립적이기에 다른 문제에 쉽게 적용 가능하다.	• 규칙이 많아지면, 규칙 간에 불명확한 상호 관계가 발생할 수 있다.
	• 인간처럼 다양한 조건 혹은 바뀐 조건에 대응하여 새로운 상황의 문제를 풀어내는, 즉 문제 해결 방법을 재조정하기가 힘들다.

■ 전문가 시스템의 역사

- 최초의 전문가 시스템은 1965년에 파이겐바움(Feigenbaum)이 개발한 덴드럴(Dendral)이라는 시스템으로서 보건 분야 인공지능인 분자의 구조를 추정하는 시스템이었다.
- 덴드럴은 최초의 실용적인 인공지능이다. 덴드럴은 원래 정치학 전공자였던 허버트 사이먼에게 지도를 받고 논리학을 전공한 학자 에드워드 파이젠바움이 유전학자 조슈아 레더버그의 외계 생명체 연구를 돕기 위해 화학 분야에서 개발한 것이다.

■ 지식 베이스

- 어떤 특수한 문제 영역에 대한 사실과 규칙으로 구성된다.
- 외부로부터 지식이 추가될 수는 있지만, 자동적으로 학습하지는 못한다.
- 추론 과정에서 기존 규칙으로부터 유도된 사실은 동적으로 추가되는 것이 가능하다.
- 전문가 시스템은 지식 베이스가 추론 기법보다 더 중요하다.

■ 지식 공학(Knowledge Engineering)

전문가로부터 문제를 풀기 위한 지식을 추출하고, 지식을 쉽게 사용할 수 있도록 체계화된 형태로 재편성하는 과정을 연구하는 학문이다.

■ 추론 엔진

- 문제를 해결하기 위해 언제, 어떤 사실들과 규칙들을 적용해야 되는지를 결정한다.
- 추론 엔진은 지식 베이스와는 상관없이 독립적이며, 어떤 전문가 시스템이라도 거의 동일한 구조를 지닌다.
- 추론 연산을 기반으로 해(solution)를 찾기 위한 탐색으로 구현된다.

2 규칙(Rule)은 IF(사실, 조건, 상황, 전제) THEN(행동, 결과, 결론)을 기반으로 한다. 자율 주행 자동차를 위한 규칙을 예시를 가지고 다음 괄호 안을 채워 보자.

Rule #1	IF 신호등이 녹색이다 THEN 자동차를 진행한다.
Rule #2	IF 신호등이 빨간색이다 (THEN 자동차를 멈춘다.)
Rule #3	IF 자동차가 움직이지 않는다 AND 연료탱크가 비었다 (THEN 자동차에 연료를 공급한다.)
Rule #4	IF 주행가능거리 < 100 km (THEN 연료경고등을 켜다.)

3 추론 엔진은 문제를 해결하기 위해 언제, 어떤 사실(Fact)들과 규칙(Rule)들을 적용해야 되는지를 결정하며, 추론 연산을 기반으로 해를 찾기 위한 탐색으로 구현한다. 동물 식별 전문가 시스템 규칙 예시를 가지고 다음 빈칸을 채워 보자.

예 동물 식별 전문가 시스템 규칙

Rule #1	IF 동물이 털이 있다 THEN 포유류이다.
Rule #2	IF 동물이 포유류이다 AND 고기를 먹는다 THEN 육식동물이다.
Rule #3	IF 동물이 육식동물이다 AND 황갈색이다 AND 갈기가 있다 THEN 사자다.
Rule #4	IF 동물이 육식동물이다 AND 황갈색이다 AND 검은 줄무늬가 있다 THEN 호랑이다.

예 알려진 사실

Fact #1	동물이 털이 있다.
Fact #2	고기를 먹는다.
Fact #3	황갈색이다.
Fact #4	갈기가 있다.

	추론으로 유추된 사실	추론에 사용된 규칙과 사실
Fact #5	포유류이다.	Rule #1, Fact #1
Fact #6	육식동물이다.	Rule #2, Fact #2, Fact #5
Fact #7	사자다.	Rule #3, Fact #3, Fact #4, Fact #6

4 역방향추론은 목표를 설정하고 추론 엔진은 이를 증명하는 증거를 찾는 방법으로, 목표를 결론으로 갖고 있는 규칙을 찾고, 이 규칙을 전제를 역으로 찾아가는 방식을 말한다. 동물 식별 전문가 시스템 규칙 예시를 가지고 동물은 사자인지 알아보고자 한다. 다음 빈칸을 채워 보자.

예 동물 식별 전문가 시스템 규칙

Rule #1	IF 동물이 털이 있다 THEN 포유류이다.
Rule #2	IF 동물이 포유류이다 AND 고기를 먹는다 THEN 육식동물이다.
Rule #3	IF 동물이 육식동물이다 AND 황갈색이다 AND 갈기가 있다 THEN 사자다.
Rule #4	IF 동물이 육식동물이다 AND 황갈색이다 AND 검은 줄무늬가 있다 THEN 호랑이다.

예 알려진 사실

Fact #1	동물이 털이 있다.
Fact #2	고기를 먹는다.
Fact #3	황갈색이다.
Fact #4	갈기가 있다.

목표: 식별한 동물은 사자입니까?

목표	추론으로 유추된 사실	추론에 사용된 규칙과 사실
사자다	동물이 육식동물이다 AND 황갈색이다 AND 갈기가 있다	Rule #3, Fact #3, Fact #4
육식동물이다	동물이 포유류이다 AND 고기를 먹는다	Rule #2, Fact #2
포유류이다	동물이 털이 있다	Rule #1, Fact #1

지도의 주안점

- 교사는 학생들이 스스로 인공지능에 대한 개념을 가질 수 있도록 관련된 개념을 소개한다.
- 인공지능이 발전되어 온 역사 속에 수학이 기여하였음을 학생 스스로 찾아볼 수 있도록 한다.
- 전문가 시스템이 과거에만 쓰였거나 한 가지 방법으로만 구현되는 것이 아니라는 것을 알 수 있게 안내한다.
- 여러 가지 추론을 통해 체계적으로 사고하는 경험을 해보며, 가장 효율적인 절차를 찾는 것이 인공지능이 작동하는 원리와 연관됨을 안내한다.

핵심 발문

- 해결해야 하는 문제의 목표는 무엇인가?
- 따라야 하는 규칙과 최적의 방법은 무엇인가?
- IF로 시작된 것과 연결되는 THEN은 무엇인가?
- 어떤 수학 개념을 알기 위해서 사전에 알아야 하는 내용들은 무엇인가?

교과서 연계 방안

- 추론은 명제 학습과 병행하여 학습할 수 있다.



관련 교과 역량

- 의사소통
 - 계산한 결과를 올바른 표현으로 설명하는가?
 - 자신의 답을 논리적으로 명확하게 설명하는가?
 - 모둠원의 의견을 경청하고 존중하는가?
- 정보 처리
 - 추론에 필요한 자료를 적절하게 선택하여 분석할 수 있는가?
 - 인공지능에 담긴 원리를 해석할 수 있는가?
 - 추론 규칙을 올바르게 활용하여 문제를 해결할 수 있는가?
- 태도 및 실천
 - 인공지능에 활용된 수학의 유용성과 가치를 인식할 수 있는가?
 - '인공지능과 수학' 활동지를 성실히 작성하고 적극적으로 참여하는가?
 - 추론에 어려움을 겪는 모둠원을 배려하고 존중하는가?

평가 결과 피드백

- 주어진 규칙을 이해하며 체계적으로 사고할 수 있는지를 확인하고, 잘못 생각한 경우 학생이 스스로 오류를 찾고 수정할 수 있게 한다.
- 여러 조건으로부터 관계를 탐구하는 문항에서 논리적으로 결과를 도출한 학생의 활동 사례를 학교 생활 기록부에 기록한다.

2차시 | 진리표와 순서도

학습 목표	인공지능에서 활용되는 진리표와 순서도의 개념을 이해하고 작성할 수 있다.
수업의 의도	수업 의도 전문가 시스템의 추론 엔진을 구성하는 수리 논리를 이해할 수 있도록 한다. 또한 여러 가지 진리표와 순서도를 작성할 수 있도록 한다. 수업의 주안점 논리 연산을 위한 기호 종류와 진리표를 이해하고, 진리표를 작성할 수 있도록 한다. 또한 규칙에 따른 순서도를 이해하고 작성해 볼 수 있도록 한다. 수업 활동 시 유의점 명제 결합 기호나 진릿값은 정해진 약속을 학습한 후 간단한 진리표를 구함으로써 논리를 자연스럽게 이해할 수 있도록 구성한다. 또한 효율적인 알고리즘 작성을 위해 순서도를 나타내는 것이 효율적임을 알게 하고 순서도 작성에 대한 예시는 학생의 수준에 따라 제시한다. 평가의 주안점 진리표와 순서도 작성 과정에서 올바른 추론을 해 낼 수 있는지 평가한다.
준비물	활동지, 자기 평가 및 동료 평가지

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	• 2~4인 1조로 모둠 구성 • 활동지, 자기 평가 및 동료 평가지 배부 • 인공지능 관련 기사 소개	
수행	활동	• 명제 논리 이해하기 • 논리 연산을 위한 기호 종류와 진리표를 확인하기 • 진리표 완성하기 • 순서도 기호 확인하기 • 순서도 알고리즘 이해하기 • 순서도 작성하기	관찰 평가
정리		• 자기 평가 및 동료 평가지 작성	자기 평가, 동료 평가



진리표와 순서도는 어떻게 작성할까?

년 월 일



전문가 시스템의 규칙을 사용해 결과를 추론하는 프로그램을 추론 엔진이라 한다. 추론 엔진은 컴퓨터에서 규칙을 해석해 처리할 때 수리논리를 이용한다. 다음의 내용을 살펴보고 다음의 물음에 답해 보자.

추론 엔진을 이용할 때 가장 기본적으로 사용하는 수리논리학의 체계는 명제 논리이다. 명제는 두 개 이상 결합하여 사용되기도 한다. 이렇게 여러 명제를 결합할 때는 논리 연산자를 이용한다. 논리 연산자에는 부정, 논리곱, 논리합, 배타적 논리합 등이 있다. 명제 논리는 명제 변수와 논리 연산자(연결 연산자)를 이용해서 논리식을 구성한다. 하지만 명제 자체의 의미를 따지지는 않는다. 단지 논리 연산자를 이용해서 명제 사이의 관련성을 나타내기만 한다. 이 결과에 술어 논리 등을 이용해 의미를 부여하면 추론 엔진으로 동작할 수 있는 논리 구성이 된다. 이러한 논리는 컴퓨터 관련 학문이나 공학 등의 분야에 폭넓게 응용되며, 특히 규칙을 기반으로 하는 인공지능에 이론적 토대가 되기도 하였다. 논리 연산은 명제(propositional) 논리와 술어(predicate) 논리로 이루어지는데, 명제 논리는 전체의 참과 거짓을 판별하는 법칙을 다루며, 술어 논리는 주어와 술어로 구분하여 참과 거짓 법칙을 다룬다.

• 활동팁 명제 논리는 인공지능 알고리즘에 꼭 필요한 개념이다. 진리표는 단순 문제나 복합 명제의 진릿값을 표로 나타낸 것이다. 명제 논리와 술어 논리에서는 기호를 사용해 문장을 표현하는데 명제 논리의 논리식은 논리 원자인 명제 변수와 논리 연산자로 구성되어 있으며, 술어 논리의 논리식은 기호를 조합해 논리 표현의 폭을 넓힌다.

1 논리 연산은 명제에서 출발한다. 명제(proposition)는 명확하게 참(T, TRUE), 거짓(F, FALSE)을 판단할 수 있는 문장을 말한다. 다음 문장이 명제인지 아닌지 판정하고 명제인 경우 참, 거짓을 말해 보자.

2×7 의 값은 홀수이다.	(정확하게 거짓이라 판별할 수 있으므로 거짓인 명제)
사과는 맛있다.	(참 또는 거짓이 구분되지 않으므로 명제가 아님.)
서울은 대한민국의 수도다.	(명확하게 참이라 판별할 수 있으므로 참인 명제)
$1 + 1 = 3$	(정확하게 거짓이라 판별할 수 있으므로 거짓인 명제)



논리 연산을 위한 기호 종류와 진리표이다. 다음의 물음에 답해 보자.

1. 진릿값(Truth Value)

진릿값이란 사실 또는 거짓을 나타내는 것으로 True(T), False(F) 또는 1, 0으로 표현한다.

2. 논리 연산자(Logical Operators)

명제는 두 개 이상 결합하여 사용되기도 한다. 이렇게 여러 명제를 결합할 때는 논리 연산자를 이용한다. 논리 연산자에는 부정(NOT, $\sim$, $\neg$), 논리곱(AND, $\wedge$), 논리합(OR, $\vee$), 배타적 논리합(Exclusive OR, XOR) 등이 있다.

3. 합성명제(Compound Proposition)

합성명제는 하나 이상의 명제들이 결합되어 만들어진 명제로 AND, OR, NOT 등의 논리 연산자를 이용해 명제들을 결합한다. 합성명제의 진릿값을 알려면 진리표(Truth Table)를 이용하면 된다.

1) 부정(Negation) NOT

문장 p 가 명제일 때 “ p 가 아니다.”도 명제이며 $\neg p$ 또는 $\sim p$ 로 표현한다. 명제 p 의 진릿값이 거짓(F)이면 명제 p 의 진릿값은 거짓(F)에 대한 부정인 참(T)이 된다. 부정 $\sim p$ 의 진리표는 오른쪽과 같다.

p	$\sim p$
T	F
F	T

2) 논리곱(Conjunction) AND

문장 p , q 가 명제일 때 p , q 의 진릿값이 모두 참(T)인 경우만 참(T)이 되고, 그렇지 않은 경우는 거짓(F)이 되는 명제를 $p \wedge q$ 로 표현한다. $p \wedge q$ 는 ‘ p and q ’ 또는 ‘ p 그리고 q ’라고 읽는다. 논리곱 $p \wedge q$ 의 진리표는 오른쪽과 같다.

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

3) 논리합(Disjunction) OR

문장 p , q 가 명제일 때 p , q 의 진릿값이 모두 거짓(F)인 경우만 거짓(F)이 되고, 그렇지 않은 경우는 참(T)이 되는 명제를 $p \vee q$ 로 표현한다. $p \vee q$ 는 ‘ p or q ’ 또는 ‘ p 또는 q ’라고 읽는다. 논리합 $p \vee q$ 의 진리표는 오른쪽과 같다.

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

4) 배타적 논리합(Exclusive OR) XOR

문장 p, q 가 명제일 때 p, q 의 진릿값 중 하나만 참(T)인 경우만 참(T)이 되고, 그렇지 않은 경우는 거짓(F)이 되는 명제를 $p \oplus q$ 로 표현한다. $p \oplus q$ 는 '배타적 논리합 q ' 또는 ' p xor q '라고 읽는다. 배타적 논리합 $p \oplus q$ 의 진리표는 오른쪽과 같다.

p	q	$p \oplus q$
T	T	F
T	F	T
F	T	T
F	F	F

5) 조건명제(Conditional Proposition) $p \rightarrow q$

문장 p, q 가 명제일 때, 명제 p 가 조건 또는 원인으로 제시되고, 명제 q 가 결론 또는 결과로 제시되는 명제를 $p \rightarrow q$ 로 표현한다. 조건명제 $p \rightarrow q$ 는 조건이 되는 명제 p 가 참(T)이고 결론이 되는 명제 q 가 거짓(F)인 경우에만 거짓(F)이 되고, 그 외의 경우에는 참(T)이 된다. 조건명제 $p \rightarrow q$ 의 진리표는 오른쪽과 같다.

p	q	$p \rightarrow q$
T	T	T
T	F	F
F	T	T
F	F	T

합성명제의 진릿값은 다음과 같이 진리표를 이용하면 된다.

p	q	$\sim p$	$p \wedge q$	$p \vee q$	$p \oplus q$	$p \rightarrow q$	$\sim p \rightarrow q$
T	T	F	T	T	F	T	T
T	F	F	F	T	T	F	T
F	T	T	F	T	T	T	T
F	F	T	F	F	F	T	F

1 진리표를 이용해서 빈칸을 채워 보자.

p	q	$\sim p$	$\sim q$	$\sim p \wedge q$	$p \vee (\sim p \wedge q)$	$\sim(p \vee (\sim p \wedge q))$	$\sim p \wedge \sim q$
T	T	F	F	F	T	F	F
T	F	F	T	F	T	F	F
F	T	T	F	T	T	F	F
F	F	T	T	F	F	T	T

2 다음의 명제들의 진릿값을 구해 보자.

사과는 채소이거나 시금치는 채소이다.	(T)
서울은 대한민국의 수도이고, $3 \times 2 = 5$ 이다.	(F)
유채꽃이 노랗다면, 바다가 육지이다.	(F)
$3 \times 2 = 5$ 라면, 런던은 미국의 수도이다.	(T)

3 명제 “정수 x 에 대해 $x \geq 50$ 이면, $x \leq 30$ 이다.”에 대해 $x=72$, $x=23$ 일 때 진릿값을 구해 보자.

$x=72$ 일 때

명제: $x \geq 50$ 이면, $x \leq 30$ 이다.

72는 50보다 크므로 명제 $x \geq 50$ 에 대해 참(T)이지만, $x \leq 30$ 에 대해 거짓(F)이다.

따라서 본 명제의 진릿값은 거짓(F)이다.

$x=23$ 일 때

명제: $x \geq 50$ 이면, $x \leq 30$ 이다.

23은 50보다 작으므로 명제 $x \geq 50$ 에 대해 거짓(F)이지만, $x \leq 30$ 에 대해 참(T)이다.

따라서 본 명제의 진릿값은 참(T)이다.

 규칙을 기반으로 하는 시스템을 만들기 위해 순서도를 어떻게 활용할 수 있는지 알아보자.

1950년대 이후 조건 분기를 활용한 컴퓨터 프로그램(응답 시스템)을 만들 수 있게 되면서 사용자가 입력한 데이터에 조건 분기를 적용하는 규칙 기반 모델이 발전하였다. 여기서 조건 분기란 어떤 특정한 조건을 비교해서 처리할 일을 나누는 것을 말하며, 컴퓨터를 이용한 문제 해결은 조건 분기를 구현한 프로그램을 실행해 답을 끌어내는 것이다. 이렇게 규칙(조건 설정)을 사용해 조건 분기 프로그램을 실행하는 시스템을 규칙 기반 시스템이라고 한다.

‘사람이 하는 일을 기계가 한다.’는 관점에서는 조건 분기 프로그램도 인공지능이라 할 수 있다. 초기의 인공지능은 조건 분기 프로그램으로 구성되었고 조건 분기 프로그램 자체는 오늘날 인공지능 외의 다양한 분야에서 사용되기도 한다. 규칙 기반 모델은 설정한 규칙에 따라 입력 데이터를 자동으로 분석하는 추론 엔진이나 규칙 설정을 외부 저장 장치 등에 쌓는 지식 기반 모델에 이용한다. 이러한 규칙 기반 모델에는 IF-THEN 형태의 알고리즘 순서도를 이용하여 규칙을 설정한다.

• **활동팁** 어떤 문제를 해결하기 위한 유한 번의 절차를 ‘알고리즘’이라고 한다. ‘순서도’는 이 알고리즘을 알기 쉽게 기호와 그림으로 나타낸 것이다. 즉, 순서도란 처리 하고자 하는 문제를 분석하여 그 처리 순서를 단계화시켜 상호간의 관계를 일정한 기호를 사용하여 일목요연하게 나타낸 그림으로 알기 쉽게 도식화하는 것을 의미한다. 순서도는 프로그래밍에서 많이 활용되고 있다.

1 선택을 해야 하는 상황에서 어떤 순서와 우선순위에 따라 규칙을 설정하게 되는지 말해 보자.

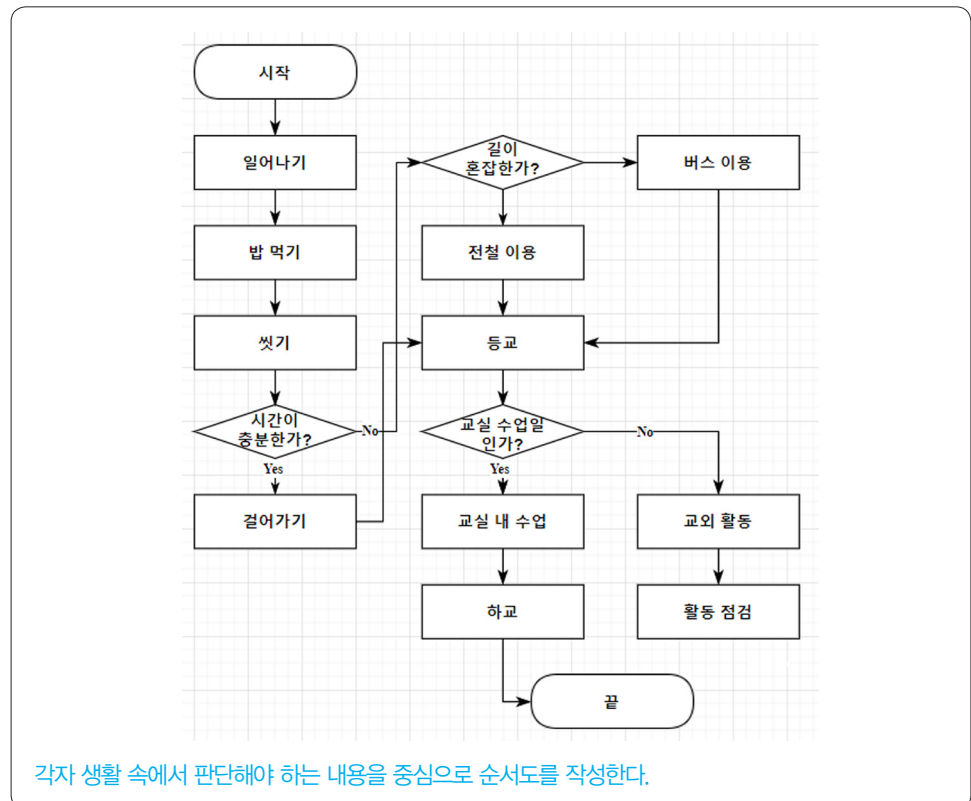
여름철 에어컨을 설정할 때, 무조건 세게만 트는 것이 아니라 에너지 절약과 건강을 생각해서 효율적으로 바람의 세기를 조절할 필요가 있다. 따라서 온도에 따른 바람의 세기를 조절하는 기준을 다양하게 세울 수 있다. 예를 들면, 높은 온도의 조건을 따져서 33도 이상이면 ‘매우 강하게’, 30도 이상이면 ‘강하게’, 28도 이상이면 ‘약하게’ 등으로 기준을 세울 수 있다.



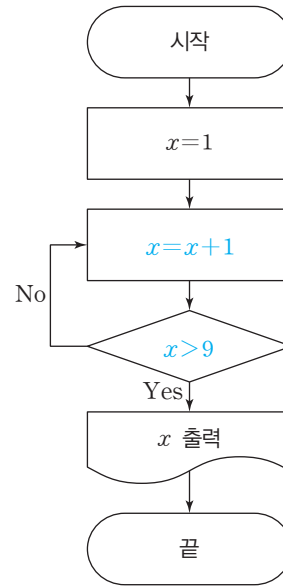
순서도(flowchart)란 어떤 문제를 해결하기 위하여 유한 번의 계산 방법 또는 처리 순서인 ‘알고리즘’을 기호와 그림으로 나타낸 것이다. 순서도에는 국제표준화기구(International Standard Organization)에서 정한 몇 가지 기호를 활용한다. 순서도에서 활용하는 기호 중 일부를 살펴보고 다음의 물음에 답해 보자.

기호	뜻	보기
	순서도의 시작 또는 끝을 나타낸다.	시작
	연산, 실행, 명령 등 모든 처리 기능을 나타낸다.	A 에 저장된 값에 2를 더하여 X 에 저장해라.
	판단 기호로서 마름모 안에 판단 기준을 쓴다.	$n=4$ 이면 ‘예’로, $n \neq 4$ 이면 ‘아니요’로 옮겨라.
	저장된 값을 인쇄하는 출력을 나타낸다.	X 에 저장된 값을 인쇄해라.

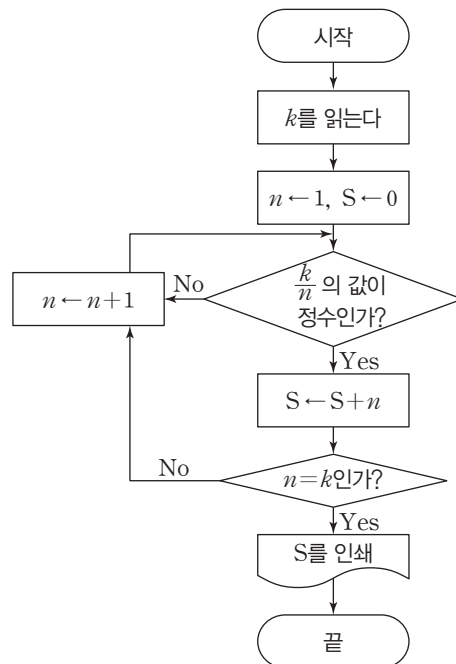
1 순서도 그리는 사이트(<https://app.diagrams.net/>)를 이용해서 하루 일과를 작성해 보자.



2 변수 X 를 1부터 시작하여 1씩 증가시켜 10이 되면 출력하는 순서도를 채워 보자.



3 다음 순서도에서 k 의 값이 12일 때 인쇄되는 S 의 값을 구해 보자.



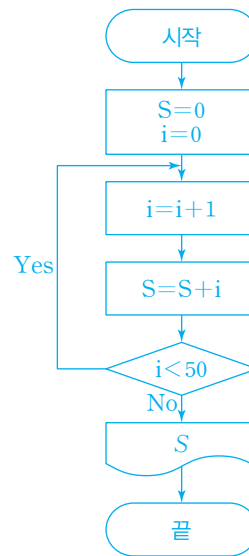
구하려는 값은 12의 양의 약수의 총합이다.

12의 양의 약수는 1, 2, 3, 4, 6, 12이므로 $S=1+2+3+4+6+12=28$ 이다.

4 알고리즘, 순서도를 컴퓨터에서 동작할 수 있도록 컴퓨터 언어로 변화하는 것을 프로그래밍 또는 코딩이라고 한다. 다음은 파이썬으로 작성한 코드이다. 아래 코드에 해당되는 순서도를 작성해 보자.

```
s, i = 0, 0
i+=1
s+=i
while i<50:
    print(s)
```

1부터 50까지 더하는 프로그래밍을 나타낸 것이다.



시도의 주안점

- 교사는 인공지능에서 활용되는 진리표와 순서도의 작성 규칙을 학생들이 이해할 수 있도록 한다.
- 논리 연산은 명제에서부터 이해할 수 있도록 한다.
- 진리표를 작성하기 위한 진릿값을 이해하고 이를 활용하여 조건 명제의 진릿값을 구할 수 있도록 한다.
- 순서도를 그리기 위한 규칙을 이해하고 이를 활용하여 생활 장면 및 간단한 수학적 문제를 해결하기 위한 순서도를 작성해 볼 수 있도록 한다.

핵심 발문

- 해결해야 하는 문제의 목표는 무엇인가?
- 따라야 하는 규칙과 최적의 방법은 무엇인가?
- 각 명제의 논리 연산자를 무엇인가? 이때의 참과 거짓은 어떻게 판단해야 하는가?
- 순서도 알고리즘에서 판단해야 할 것이 무엇인가? 또 판단에 따라 어디로 가야 하는가?

교과서 연계 방안

- 진리표는 명제 학습과 병행하여 학습할 수 있다.



관련 교과 역량

- 의사소통
 - 계산한 결과를 올바른 표현으로 설명하는가?
 - 자신의 답을 논리적으로 명확하게 설명하는가?
 - 모둠원의 의견을 경청하고 존중하는가?
- 정보 처리
 - 명제의 참, 거짓을 판단할 수 있는가?
 - P와 Q의 진리표에 진릿값을 결정할 수 있는가?
 - 순서도의 알고리즘에 따라 문제를 해결할 수 있는가?
 - 주어진 코드를 해석하고 이를 순서도로 변환할 수 있는가?
- 태도 및 실천
 - 인공지능에 활용된 수학의 유용성과 가치를 인식할 수 있는가?
 - '인공지능과 수학' 활동지를 성실히 작성하고 적극적으로 참여하는가?
 - 추론에 어려움을 겪는 모둠원을 배려하고 존중하는가?

평가 결과 피드백

- 진리표와 순서도의 규칙을 이해하며 체계적으로 사고할 수 있는지를 확인하고, 잘못 생각한 경우 학생이 스스로 오류를 찾고 수정할 수 있게 한다.
- 여러 조건으로부터 관계를 탐구하는 문항에서 논리적으로 결과를 도출한 학생의 활동 사례를 학교 생활 기록부에 기록한다.

3차시 | 생활 속의 인공지능과 수학

학습 목표	• 우리 생활 주변에 활용되는 인공지능을 찾아볼 수 있다. • 생활 속의 인공 지능에 어떤 수학적 원리가 활용되는지 이해할 수 있다.
수업의 의도	수업 의도 추천 시스템, 영상 분류, 날씨 예측 등 우리 주변에서 활용되고 있는 인공지능의 사례를 직접 찾아보고 이러한 기능들의 특징을 이야기할 수 있도록 한다. 또한 인공지능의 기능 중 기계학습과 딥러닝, 지도학습, 비지도학습의 특징을 살펴 보고 인공지능에서 활용되는 분류와 회귀, 클러스터링에 활용되는 수학 개념을 이해함으로써 수학의 가치와 유용성을 인식할 수 있게 한다. 수업의 주안점 ‘스팸 메일과 정상 메일을 분류’하거나 ‘판매액, 원유, 주택 가격의 예측’ 등에서 인공지능이 활용하는 분류나 회귀 알고리즘의 예시와 수학적 개념을 연결 지을 수 있도록 한다. 학생들이 인터넷 등을 활용하여 인공지능이 활용되고 있는 사례를 직접 찾아보면서 의사소통할 수 있게 한다. 수업 활동 시 유의점 인공지능이 학습을 통해 판단과 예측을 할 수 있는 기계학습 등을 소개하는 장면에서는 학생들의 흥미와 이해 수준을 반영하여 내용과 범위를 정한다. 학생들이 자기 주도적으로 정보를 찾고 모둠별로 활동하는 과정에서 서로 관심사와 이해 정도를 공유할 수 있게 한다. 평가의 주안점 분류나 회귀에 사용되는 수학적 개념을 이해하고 적용할 수 있는지, 기계학습을 경험하는 과정에서 자신의 의견을 잘 표현할 수 있는지 평가한다.
준비물	활동지, 인터넷(단말기), 자기 평가 및 동료 평가

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	• 2~4인 1조로 모둠 구성 • 활동지, 자기 평가 및 동료 평가지 배부 • 우리 생활 주변에 활용되고 있는 인공지능 소개	인공지능 활용 예를 찾는 과정에서의 정의적 평가
수행	활동	• 생활 속 활용되는 다양한 인공지능 찾아보기 • 기계학습이 무엇인지 알아보고 경험하기 • 함수 $y=f(x)$에서 변수 y에 해당되는 것 찾아보기 • 학습의 방법인 지도학습과 비지도학습 알아보기 • 분류나 회귀 알고리즘을 알아보기 • 분류나 회귀에 활용되는 수학적 개념 알아보기	관찰 평가
정리		• 자기 평가 및 동료 평가지 작성	자기 평가, 동료 평가

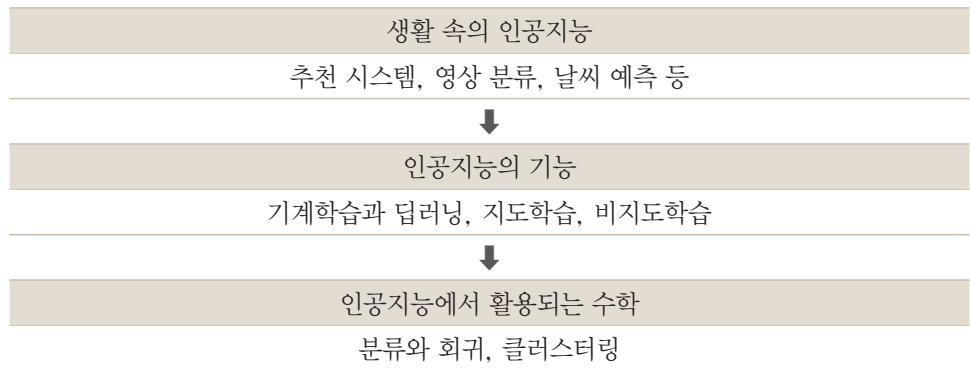


생활 속 인공지능에서 활용되는 수학이야기

년 월 일



생활 속의 인공지능에 어떤 수학적 원리가 활용되는지 알아보자.

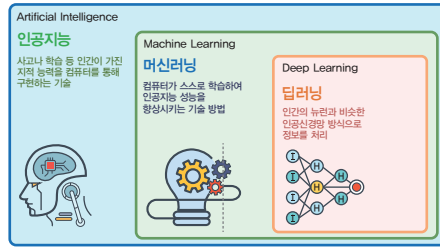


우리 생활 주변에는 다양한 인공지능이 활용되고 있다. 인터넷 등을 활용하여 인공지능이 활용되고 있는 사례를 찾아보자.

우리 생활주변에 활용되는 인공지능의 예는 Siri, Google Assistant 등과 같은 음성 어시스턴트와 LINE이나 메신저 등 채팅 도구를 이용하여 인공지능이 인간의 질문에 응답하는 기능인 챗봇이 있다. 또한 유튜브(YouTube), 넷플릭스(Netflix)의 추천 알고리즘과 사이트 이용 및 상품 이력 조화를 기반으로 “추천 상품”이라고 말하고 있는 AI, 방에 떨어진 쓰레기의 양에 맞추어 흡입력을 조정하여 청소하는 로봇 청소기나 취향에 맞게 밥알의 무르기를 조절하는 전기밥솥 등과 같은 “스마트 가전”, 대화를 하는 것뿐만 아니라 오늘의 날씨, 개인 스케줄을 알려주거나, 음식을 배달해 주고, 상품을 구입할 수 있는 “스마트 스피커”, 호텔의接客 로봇, 아이들의 장난감에 AI를 탑재하여 아이들의 놀이 상대로의 로봇 등으로 사용되는 “서비스 로봇”, 사람이 운전하지 않아도 자동차가 자율적으로 목적지에 도착하는 “자율주행 시스템”, 이미지 데이터에서 병명을 진단하거나 논문 데이터를 학습하여 병명을 알아내는 의료 분야에서 활약하고 있는 “AI 닥터” 선수나 경기 관련 데이터 분석을 도와주고, 오심을 줄이기 위한 경기의 판정을 도와주는 등 스포츠계에서 활용되는 AI 등이 있다. SNS가 활성화된 요즘 마케팅 또한 매우 중요하데, 마케팅으로써 활용되는 인공지능은 홈페이지에 들어오는 방문자의 행동과 패턴을 분석하여 더 나은 홈페이지의 개선방향을 제안하는 AI와 현재까지의 웹사이트를 조회한 이력을 바탕으로 관심이 있다고 판단하여 보내고 싶은 대상에게만 광고를 전달하는 AI가 있다.



인공지능이 학습을 통해 판단과 예측을 할 수 있는 기계학습(Machine Learning)에 대한 설명을 살펴보자.

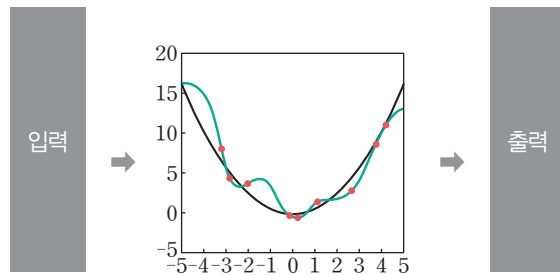


지식기반 방식에서 기계학습으로의 인공지능 기술 대전환

기계학습은 사람이 학습하듯이 컴퓨터가 데이터를 가지고 학습하게 함으로써 새로운 지식을 얻을 수 있도록 하는 분야로, 규칙기반 프로그래밍에 의존하지 않고 데이터로부터 학습할 수 있는 알고리즘을 기반으로 한다.

■ 학습(learning)

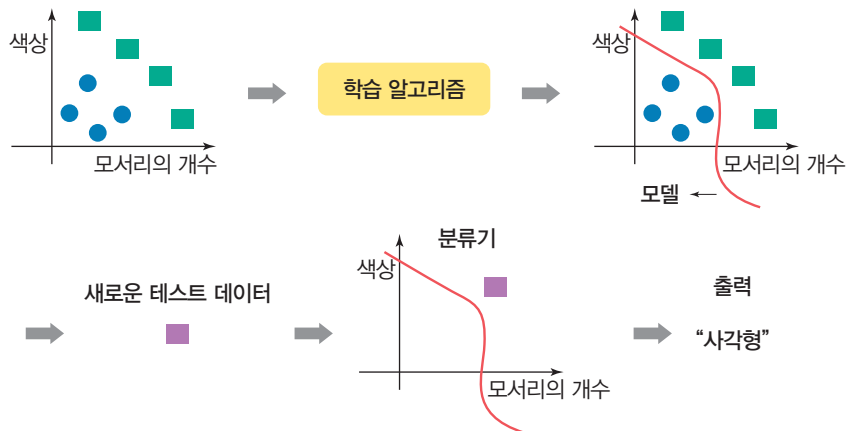
- 모델을 만들거나 배우는 것을 의미한다.
- 기계학습은 항상 입력을 받아서 출력하는 함수 $y=f(x)$ 를 학습한 후 예측한다. 즉, 함수를 이용한 학습이다.



기계학습(machine learning) ≈ 함수 근사(function approximation)

■ 예측(prediction)

- 학습된 모델을 레이블이 없는 샘플에 적용하는 것을 의미한다.
 - 즉, 학습된 모델을 사용하여 유용한 예측(y)을 해내는 것이다.
- ※ 기계학습에 주어지는 데이터들에 y 값(레이블, label)이 있는 데이터도 있고 없는 데이터도 있다.



1 기계학습에서 활용되는 함수 $y=f(x)$ 에서 변수 y 에 해당되는 것들이 무엇이 있을 수 있는지 생각하여 답해 보자.

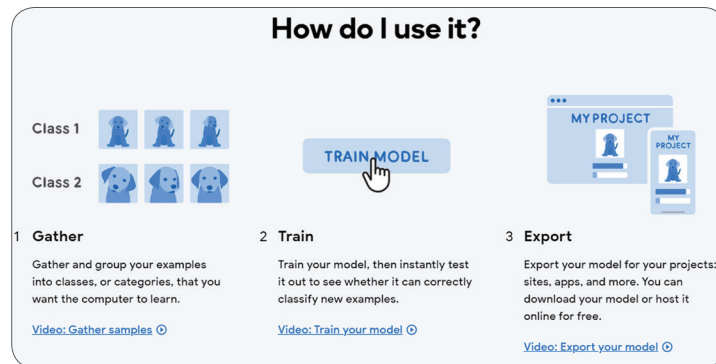
데이터(샘플)	변수 y
농작물 가격	농작물의 향후 가격
동물 사진	사진에 표시되는 동물의 종류
동영상	선호하는 동영상
기상 일지	장마 날짜

2 인터넷 등을 활용하여 인공지능에서 기계학습을 활용하는 사례를 찾아보자.

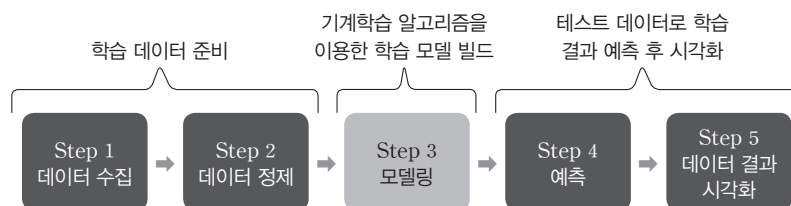
- 객체 인식: 개와 고양이 이미지 인식-코, 귀, 크기 등 객체의 특징을 구분하여 인식하고 새로운 데이터 (테스트 데이터)를 입력하면 이 기준에 따라 개인지 고양이인지 분류한다.
- 야외 경기 날짜 예측: 테스트 경기를 진행했던 날씨 조건을 학습하여, 특정 날씨에 경기를 진행할 수 있을 지를 예측하는 문제이다.
- 필기 문자인식: 다양한 필기 문자에 대한 데이터를 학습한 후 특정 필기에 대한 문자를 인식하는 문제이다.

보충 자료

- 기계학습 경험하기-Google Teachable Machine을 활용하여 기계학습을 경험해 보자.
- Teachable Machine(<https://teachablemachine.withgoogle.com>)



- 구글에서 제공하는 기계학습을 학습하기 위한 서비스이다.
- 이미지, 소리, 자세(pose) 인식과 학습 기능을 제공한다.
- 인공지능이나 기계학습에 대한 사전 지식과 코딩 능력이 없이도 활용이 가능하다.
- 별도의 프로그램 설치 없이 크롬 브라우저에서 Teachable Machine 홈페이지에 접속해서 바로 사용할 수 있다.
- 다음의 과정을 통해 기계학습을 수행한다.





기계학습 안에는 지도학습과 비지도학습 방법이 있다. 다음의 물음에 답해 보자.

학습의 방법인 지도학습과 비지도학습

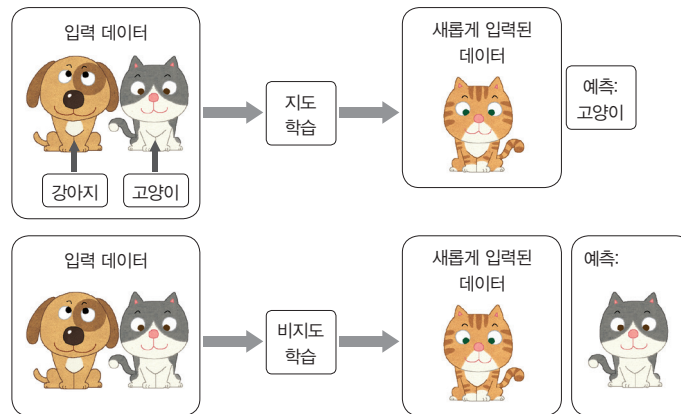
지도학습은 문제와 답이 함께 주어진 데이터를 학습하는 알고리즘이다. 컴퓨터는 주어진 문제에 대한 답을 예측하기 위한 모델을 찾는다. 답의 형태에 따라서 회귀분석과 분류 분석으로 나눌 수 있다. 비지도학습은 답이 주어지지 않은 문제만 주어진 데이터를 학습하는 알고리즘이다. 컴퓨터는 문제 속에 숨은 규칙성을 스스로 찾는다.

지도학습 - 문제와 정답을 모두 알려주고 공부시키는 방법

비지도학습 - 답을 가르쳐주지 않고 공부시키는 방법

강아지와 고양이 분류 문제를 예를 들면, 강아지와 고양이 이미지에 레이블까지 함께 주어진 데이터를 컴퓨터가 학습하는 것이 지도학습이다. 컴퓨터는 이미지에 대한 레이블(label) 또는 태그(tag)를 예측하기 위하여 모델을 만든다. 새로운 고양이 이미지를 입력하면 컴퓨터는 모델을 통해서 레이블이 '고양이'임을 예측한다.

한편, 강아지와 고양이 이미지만 주어진 데이터를 컴퓨터가 학습하는 것이 비지도학습이다. 컴퓨터는 이미지를 분류하는 모델(기준)을 스스로 만든다. 새로운 고양이 이미지를 입력하면 컴퓨터는 스스로 만든 분류 모델(기준)을 통해서 '고양이'로 분류한다.



1 데이터 입력에 대한 출력(답)을 하나의 쌍으로 학습시키는 방법인 지도학습의 장점과 단점이 무엇인지 생각해 보자.

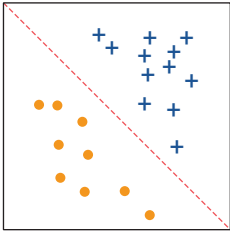
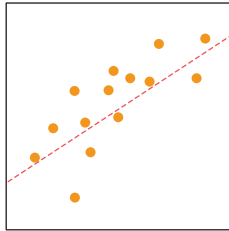
장점	단점
• 이전에 학습된 경험으로부터 데이터 출력을 생성한다. • 답(레이블)을 함께 학습시키기 때문에 정확성이 높다. • 경험을 사용하여 성능 기준을 최적화하는데 용이하다. • 다양한 유형의 문제 해결에 도움이 될 수 있다.	• 반드시 레이블이 있는 데이터를 사용해야 한다. • 빅데이터를 다루는 경우 학습을 위한 데이터를 준비하는 단계에 많은 시간이 소요될 수 있다.

2 출력값을 알려주지 않고 스스로 모델을 구축하여 학습하는 비지도학습의 장점과 단점이 무엇인지 생각해 보자.

장점	단점
- 비정제된 데이터를 입력하여 데이터의 특징을 잡아 군집(Clustering)할 수 있다. - 레이블을 제공할 필요가 없으므로 레이블을 제공하는 시간이 소요되지 않는다. - 목표값을 정해지지 않았기 때문에 레이블에 없는 새로운 결과가 창출될 가능성이 있다.	- 레이블이 없으므로 모델 성능을 평가하는 데에 어려움이 있다. - 기준을 스스로 찾아야 하므로 어렵고 정확성이 떨어진다. - 학습된 데이터의 구분하는 기준에 따라 결과가 다소 달라질 수 있다.

3 우리 생활 주변에서 ‘스팸 메일과 정상 메일을 분류’하거나 ‘판매액, 원유, 주택 가격의 예측’ 등에 인공지능을 활용하고 있다. 이러한 인공지능의 학습 방법 중 지도학습에는 분류나 회귀 알고리즘을 사용한다.

1) 분류와 회귀가 무엇인지 확인해 보자.

분류(Classification)	회귀(Regression)
- 일정한 기준에 따라 명백하게 구분 짓는 것 - 분류는 유사한 특성을 가진 데이터끼리 묶어서 그룹 짓는 것으로써 미리 정의된 가능성 있는 여러 클래스 레이블(class label) 중 어느 클래스에 속하는 지를 예측하는 것	- 회귀식을 통해 값을 예측하는 것 - 회귀는 학습한 데이터를 근거로 그 관계를 설명하는 수학적식(추세선/회귀식)을 설정하고, 이를 통해 새로 입력된 데이터에 대한 출력값을 통계적 계산을 통해 예측하는 것
	

2) 다음 사항을 예측을 하고자 할 때, 각 예측을 분류나 회귀로 구분해 보자.

“내일 날씨는 덥다/춥다.”와 같은 이분법적 선택을 한다.	분류
“내일 기온은 18.3도”와 같이 연속값을 통해 날씨를 추정한다.	회귀
(공부 시간/게임 시간, 합격 여부)의 데이터로 학습 후, (공부 시간/게임 시간) 입력으로 [합격 여부(합격/불합격)]을 예측한다.	분류
(공부 시간/게임 시간, 시험 점수)의 데이터로 학습 후 (공부 시간/게임 시간) 입력으로 [시험 점수(000점)]을 예측한다.	회귀



지도학습의 분류 알고리즘 중에서 나이브 베이즈와 K-NN에 대한 내용이다. 다음의 물음에 답해 보자.

나이브
베이즈
(Naïve
Bayesian)

나이브 베이즈 분류는 특성들 사이의 독립을 가정하는 베이즈 정리를 적용한 확률 분류의 일종이다.

즉, 측정 데이터가 여러 가지 속성(feature)을 가지고 있을 때, 어느 분류에 속할 지 판단하는 조건부확률 모델

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

을 이용한 분류 알고리즘이다.

※ 두 사건 A, B 에 대하여 사건 A 가 일어났다는 조건 아래에서 사건 B 가 일어날 확률을 사건 A 가 일어났을 때의 사건 B 의 조건부확률이라고 하고, 이것을 기호로 $P(B|A)$ 와 같이 나타낸다.

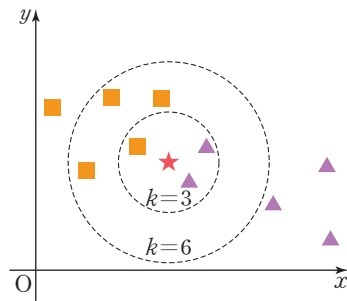
이때 $P(B|A)$ 는

$$P(B|A) = \frac{n(A \cap B)}{n(A)} = \frac{P(A \cap B)}{P(A)}$$

가 된다.

K-NN
(k -nearest
neighbors)

K-NN은 '최근접 이웃 분류'라고 불리며, 각 학습 데이터는 자기가 속하는 레이블을 가지는 분류기 역할을 수행한다.



• 가장 가까운 것들과의 거리 계산으로 범주를 분류하는 가장 간단한 기계학습 알고리즘이다.

• 측정 데이터 주위의 K개의 가장 가까운 이웃 데이터를 선택하여 추론하는데, 이때 거리 개념이 적용된다.

- 거리 계산하기
- 가장 근처에 있는 요소 뽑기
- 예측하기

※ 두 점 $A(x_1, y_1), B(x_2, y_2)$ 사이의 거리는

$$AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

이다.

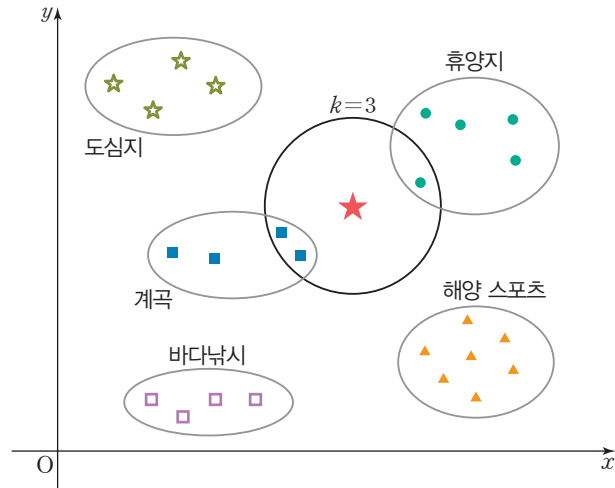
1 서울과 제주 간을 운항하는 어느 항공사의 비행기가 정시에 출발했던 데이터가 A , 정시에 도착했던 데이터를 B 라고 하면 $P(A)=0.9$, $P(A \cap B)=0.8$ 이라고 한다. 이 비행기가 정시에 출발하였다고 할 때, 정시에 도착할 확률을 예측해 보자.

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.8}{0.9} = \frac{8}{9}$$

2 세 개의 데이터 $A(-1, 1)$, $B(1, 4)$, $C(2, -1)$ 에서 두 데이터 A 와 B , B 와 C 거리를 각각 구해 보자.

$$\begin{aligned} \overline{AB} &= \sqrt{(1 - (-1))^2 + (4 - 1)^2} = \sqrt{13} \\ \overline{BC} &= \sqrt{(2 - 1)^2 + \{(-1) - 4\}^2} = \sqrt{26} \end{aligned}$$

3 휴가철 여행지를 선택하는 데 있어서 나이 x 와 장소 y 로 학습시킨 후 범주로 나타낸 그림이다. 추론 단계에서 새 데이터를 나이와 장소에 맞게 위치시킨 후 $k=3$ 으로 한 경우 새 데이터는 어떤 범주에 속한다고 추론할 수 있는지 말해 보자.



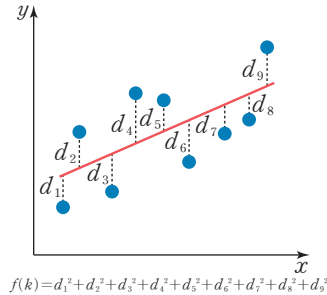
새 데이터에서 가까운 두 개의 데이터가 '계곡' 카테고리에 속하고, 1개의 데이터가 '휴양지' 카테고리에 분류된 걸 보아 이 새로운 데이터는 '계곡' 카테고리에 속한다고 추론할 수 있다.



지도학습의 회귀 알고리즘이다. 다음의 물음에 답해 보자.

회귀(Regression)는 데이터 변수들의 사이의 관계를 결정하는 통계적 추정 방법으로써 학습한 데이터를 근거로 그 관계를 설명하는 수학적(추세선)을 설정하고, 이를 통해 새로 입력된 데이터에 대한 출력값을 통계적 계산을 통해 예측하는 지도학습 기법이다.

데이터를 가장 잘 설명할 수 있는 직선 형태의 추세선은 각 점에서 회귀 직선까지의 y 축 방향의 거리 d 제곱의 총합을 최소로 해서 얻어지는 직선으로 $y=ax+b$ 로 나타낸다.



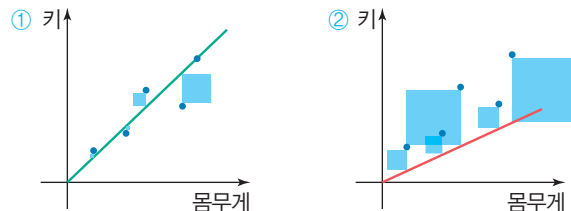
※ y 값의 차이를 구한 후 그 차이의 합이 가장 작은 값을 찾는 과정에서 음수가 나오는 것을 방지하기 위해서 절댓값(||)을 붙이는 방법도 있으나, 일반적으로 제곱을 하면 계산이 더 편리하다.

1 다음은 훈련 데이터로 몸무게로 키를 예측하는 추세선을 구현하였을 때, 어느 추세선이 더 예측에 적합한지 판단하고 그 이유를 말해 보자.



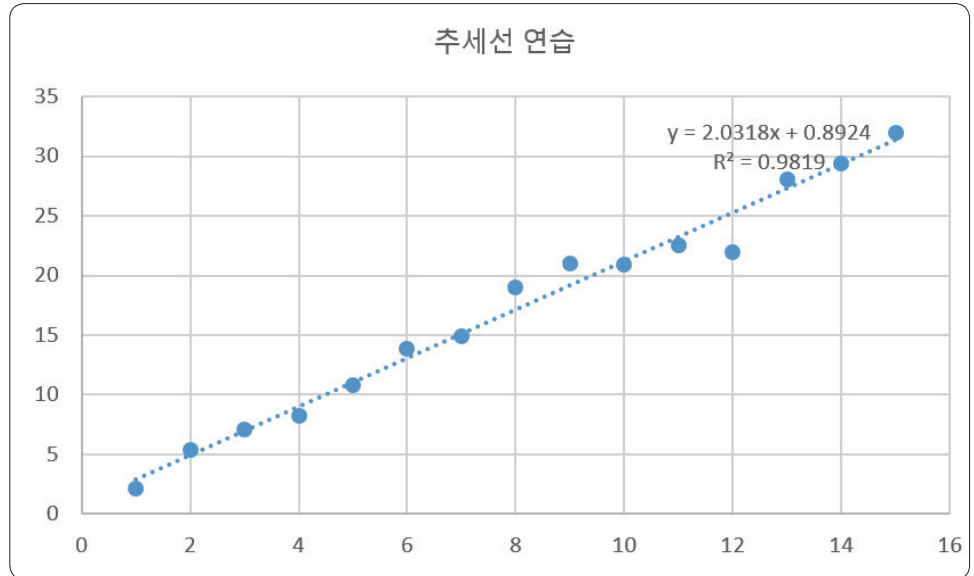
①번 추세선이 추세선과 훈련 데이터와의 오차가 더 적어 보인다.

훈련 데이터의 실제 y 값과 추세선에 의하여 예측된 y 값 사이의 차이를 제공해서 넓이로 보는 방법이 있다. 이는 모델을 평가하기 위해 오차를 제공하면 그 값이 증폭되어 그 차이를 쉽게 비교할 수 있다는 장점도 있다. 다음 그림처럼 추세선과의 차이를 제공해서 시각형을 만들어 그 합을 구해보면 ①번 그래프의 값이 더 작아 더 정확한 추세선이라 할 수 있다.



2 다음 데이터들을 스프레드시트(엑셀)을 이용해서 데이터에 적합한 추세선을 구해 보자.

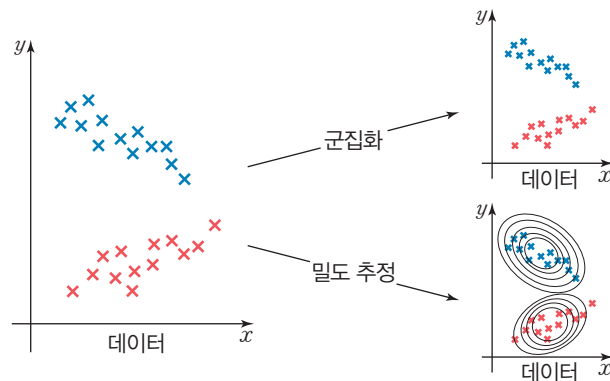
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2.1	5.4	7.1	8.2	10.8	13.8	14.9	19	21	20.9	22.5	22	28.1	29.4	32



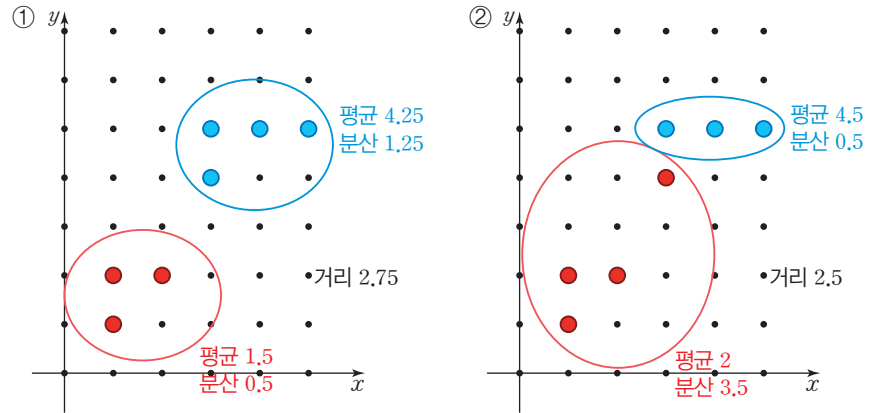
우리 생활 주변에서 ‘고객의 구매 이력으로 고객 분류’하거나 ‘질병과 치료를 위한 패턴 탐지’, ‘신용카드 사기 탐지’, ‘폭풍을 예측하는 기상 예보’ 등에서 인공지능을 활용하고 있다. 이러한 인공지능의 기술 중 비지도학습의 학습 유형의 군집화와 밀도 추정에 대한 내용을 확인해 보자.

데이터 분류에 대한 정보가 전혀 없이 패턴을 찾거나 데이터를 분류하려고 할 때 사용하는 학습 방법으로 데이터들의 상관도를 분석하여 유사한 데이터들을 모을 수 있다.
 군집화: 유사성에 따라 데이터를 분할하는 것으로 군집 내의 분산과 군집 간의 거리를 계산하는 것

밀도 추정: 부류(class)별 데이터를 만들어 냈을 것으로 추정되는 확률분포를 찾는 것



1 다음은 비지도학습을 통해 데이터를 군집화한 결과를 나타내는 그림이다. 두 경우 중 어느 군집화가 더 적합한지 판단하고 그 이유를 이야기해 보자.



위 그림과 같이 서로 다른 분산과 거리로 군집화한 경우, 두 번째 경우가 분산은 더 커지고 거리는 더 짧아진다. 일반적으로 군집 내의 분산은 작고 군집 간의 거리가 더 큰 경우가 군집 간의 차이를 더 분명히 드러낸 것이라 할 수 있다. 따라서 ①번 경우가 군집화가 더 잘 되었다고 판단할 수 있다.

지도의 주안점

- 인터넷 및 각종 자료를 통해 우리 생활 주변에는 활용되는 인공지능을 찾아볼 수 있도록 한다.
- 기계학습, 지도학습, 비지도학습이 무엇인지 알아보는 과정은 학생의 흥미와 수준에 맞도록 그 깊이와 순서를 조정한다.
- 분류나 회귀에 활용되는 수학적 개념은 간단한 문제를 적용해서, 학생들이 교과서에서 배우고 있는 수학이 관련되어 있음을 이해할 수 있도록 한다.

핵심 발문

- 함수 $y=f(x)$ 에서 변수 y 에 해당되는 것은 무엇인가?
- 지도학습, 비지도학습에서 다루는 데이터의 특징은 무엇인가?
- 분류나 회귀의 차이는 무엇이며 어떤 자료를 다룰 때 효과적인가?
- 기계학습을 활용하는 사례들에서 나타나는 특성은 무엇인가?

교과서 연계 방안

- 분류 알고리즘과 연계된 수학은 조건부확률, 두 점 사이의 거리 학습과 병행하여 학습할 수 있다. 회귀 알고리즘은 함수나 미적분의 극대, 극소와 병행하여 학습할 수 있다.



관련 교과 역량

- 의사소통
 - 안내한 앱을 통해 경험한 기계학습의 특징을 설명할 수 있는가?
 - 자신의 답을 논리적으로 명확하게 설명하는가?
 - 생활 속의 인공지능의 예를 설명할 수 있는가?
- 정보 처리
 - 인공지능에 활용되는 함수 $y=f(x)$ 에서 변수 y 에 해당되는 것을 판단할 수 있는가?
 - 지도학습과 비지도학습의 특징으로부터 장점과 단점을 결정할 수 있는가?
 - 분류나 회귀 알고리즘의 특징으로부터 사례를 구분할 수 있는가?
- 문제해결
 - 확률값을 예측하기 위한 조건부 문제를 해결할 수 있는가?
 - 데이터 간의 거리를 구할 수 있는가?
 - 분산 개념을 활용하여 적절한 클러스터링을 판단할 수 있는가?
- 추론
 - 데이터 간의 범주를 추론할 수 있는가?
 - 어느 추세선이 더 예측에 적합한지 판단할 수 있는가?

평가 결과 피드백

- 기계학습, 지도학습과 비지도학습의 특징을 이해하며 이를 자신의 말로 설명할 수 있는지를 확인하고, 잘못 생각한 경우 학생이 스스로 오류를 찾고 수정할 수 있게 한다.
- 여러 조건으로부터 관계를 탐구하는 문항에서 논리적으로 결과를 도출한 학생의 활동 사례를 학교 생활 기록부에 기록한다.

II

자료의 표현

1. 영역의 개관

수와 수학 기호는 자료를 효과적으로 표현할 수 있는 도구이고, 인공지능이 다루는 자료는 수학을 이용하여 표현된다. 인공지능이 해결하려는 문제에 주로 등장하는 텍스트 자료나 이미지 자료를 벡터나 행렬로 표현하고 연산하는 과정을 통해 수학적 문제 해결 능력을 기를 수 있게 한다.

가 2015 개정 수학과 교육과정 내용

1. 성취기준 및 학습 요소

자료의 표현 영역에 제시된 성취기준 및 학습 요소는 다음과 같다.

① 텍스트 자료의 표현

[12인수02-01] 수와 수학 기호를 이용하여 실생활의 텍스트 자료를 목적에 알맞게 표현할 수 있다.

[12인수02-02] 수와 수학 기호로 표현된 텍스트 자료를 처리하는 수학 원리를 이해하고 자료를 시각화할 수 있다.

② 이미지 자료의 표현

[12인수02-03] 수와 수학 기호를 이용하여 실생활의 이미지 자료를 목적에 알맞게 표현할 수 있다.

[12인수02-04] 수와 수학 기호로 표현된 이미지 자료를 처리하는 수학 원리를 이해한다.

관련 학습 요소

• 벡터, 행렬

각각의 학습 요소를 간단히 설명하면 다음과 같다.

벡터

자료 여러 개를 순서쌍의 형태로 표현한 것을 벡터라고 한다. 이때 각각의 자료를 벡터를 구성하는 성분이라고 한다. $a_1, a_2, \dots, a_n$ 의 n 개의 자료를 성분으로 하는 벡터는 $(a_1, a_2, \dots, a_n)$ 과 같이 나타낸다.

인공지능에서 벡터는 자료를 표현하는 방법이며 벡터로 표현함으로써 텍스트 자료를 컴퓨터가 인식할 수 있게 되고, 벡터의 내적이나 간단한 연산을 통해 문서 사이의 유사도를 구해 보거나 분류하는 등의 분석을 할 수 있다.

행렬

수를 행과 열을 가진 직사각형에 배열하여 표현한 것을 행렬이라고 한다. 자료를 성분으로 하는 벡터를 나열하여 만든 직사각형으로 생각할 수 있으며 행렬의 가로 줄을 행, 세로 줄을 열이라고 한다. 즉, m 개의 행과 n 개의 열이 있는 경우 행렬은 오른쪽 그림과 같이 나타낼 수 있다.

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & & & \\ \vdots & & & \\ a_{m1} & & & a_{mn} \end{pmatrix}$$

$[m \times n \text{행렬}]$

인공지능에서 사진 인식, 생체 인식 식별을 위한 이미지 자료를 가로와 세로로 잘게 분할하여 격자(pixel)로 나눈 다음 각 격자를 색을 나타내는 수나 행렬을 이용해 표현할 수 있으며 행렬의 덧셈이나 곱셈과 같은 연산을 통해 이미지를 변형할 수 있다.

2. 교수·학습 방법 및 유의 사항

자료의 표현 영역의 교수·학습 방법 및 유의 사항은 다음과 같다.

- 실생활의 텍스트 자료로 댓글이나 감상평 등을 다룰 수 있다.

실생활에서 쉽게 접하는 포털 사이트의 영화 감상 댓글, 배달 앱의 음식 리뷰, 쇼핑몰의 구입 후기와 같은 대상에 대한 선호도를 표현하는 글 모음이나 다양한 SNS, 게시판, 뉴스에 게시된 여러 가지 주제의 글과 기사 자료 등을 선호도 판단이나 주제 찾기 등의 목적을 가지고 분석할 수 있는 텍스트 자료의 예시로 사용할 수 있다.

- 자료는 정형, 비정형, 범주형, 연속형 등 여러 가지 유형을 다룰 수 있다.

나이, 성별, 이름과 같이 범주가 몇 가지 이름으로 이미 분류된 자료를 정형 자료라고 하며 문서, 댓글, 설명서와 같이 구문이나 문장을 그대로 포함한 텍스트 자료를 비정형 자료라고 한다. 또한 자료가 성별과 같은 범주형 자료와 몸무게와 같은 연속형 자료 등 여러 가지 형태의 자료가 있다.

- 텍스트 자료나 이미지 자료를 수와 수학 기호로 표현할 때 공학적 도구를 이용할 수 있다.

* 임베딩 벡터

단어의 빈도수로 나타내는 벡터는 차원이 매우 커지는 단점을 보완하기 위해 일반적으로 단어의 차원을 줄이되 단어의 특성이나 유사성을 가지도록 변환시킨 벡터

단어별로 단어 임베딩 벡터를 이용하여 나타내는 경우, 문장에 있는 단어들의 임베딩 벡터의 각 성분별 평균을 내어볼 수도 있다. 이때 컴퓨터 프로그램이나 다양한 코딩 프로그램을 이용하여 계산할 수 있다.

- 실생활의 텍스트 자료를 수와 수학 기호를 이용하여 다양한 방식으로 표현해 보고 토론하는 활동을 통해 각 표현 방식의 장단점을 이해하게 한다.

텍스트 자료를 집합으로 표현하거나, 등장하는 단어의 빈도수 등을 세어 벡터로 표현할 수 있다. 이때, 집합으로 나타내는 경우 단어의 빈도수는 알 수 없으며 벡터로 나타내는 경우 단어의 배열 순서 등을 알지 못한다. 이와 같이 여러 가지 표현 방식의 장단점에 대해 생각해 볼 수 있다.

- 텍스트 자료는 벡터, 이미지 자료는 행렬로 표현할 수 있음을 간략하게 소개할 수 있다.

텍스트 자료는 다양한 목적에 알맞게 단어의 빈도수 등을 세어 벡터로 나타낼 수 있으며, 이미지 자료는 작은 격자 형태로 나눈 픽셀에 색이나 밝기를 나타내는 수치를 이용하여 행렬로 표현함으로써 컴퓨터가 자료를 인식할 수 있다.

- 텍스트 자료를 변형하는 활동을 통해 자료 처리의 원리를 이해하게 하고, 이때 공학적 도구를 이용할 수 있다.

텍스트 자료의 시각화를 위해 텍스트 자료를 원그래프, 막대그래프 등 다양한 그래프 표현을 이용하여 나타내는 활동을 해 볼 수 있다. 또한 단어 빈도수의 벡터로부터 단어 구름을 그려 주제어를 찾는 시각화 과정에서 컴퓨터 프로그램이나 다양한 코딩 프로그램을 이용할 수 있다.

- 이미지 자료는 각 픽셀의 위치를 나타내는 가로, 세로 좌표와 색깔을 나타내는 정보로 구성됨을 이해하게 한다.

이미지 자료는 각 픽셀을 가로와 세로로 구분된 행렬로 표현한다. 행과 열이 이미지의 위치 정보가 되며, 숫자가 색깔을 나타냄으로써 이미지를 표현할 수 있다.

- 이미지의 구도, 색상, 휘도, 밝기, 선명도 등을 간단한 연산을 통해 변경하는 활동을 할 때, 공학적 도구를 이용할 수 있다.

이미지 자료를 수, 벡터 또는 행렬로 표현하면 수, 벡터, 행렬의 연산을 이용하여 목적에 맞게 변형하여 이미지의 구도, 물체의 형태와 위치, 색상, 휘도, 밝기, 선명도 등을 변화시킬 수 있다. 이미지 확대, 모자이크 처리, 컬러 이미지의 흑백 이미지 변환 등 실생활에서 많이 접하는 이미지 변형 활동을 위해 확대/축소, 회전, 평행이동을 나타내는 행렬과 색상, 휘도, 밝기, 선명도를 변화하는 행렬을 곱하여 그 결과를 확인하는 활동에서 공학적 도구를 활용할 수 있다.

3. 평가 방법 및 유의 사항

자료의 표현 영역의 평가 방법 및 유의 사항은 다음과 같다.

- 자료의 표현에 대한 평가는 실제 자료를 목적에 맞게 수와 수학 기호로 표현하고 처리할 수 있는지를 어렵지 않은 수준에서 간단히 다룬다.

텍스트 자료를 수나 벡터로 표현하는 다양한 방법 중에서 자료를 처리하는 목적에 맞는 방법을 적절히 선택하거나 다루는 데 중점을 둔다. 예를 들어, 단어 구름과 같은 시각화 표현을 위해 단어 빈도수 벡터가 필요함을 이해하게 한다. 한편, 임베딩 벡터를 다루는 경우 복잡한 처리 과정보다는 의미를 이해하는 수준에서만 간단히 다룬다.

2. 모듈②

텍스트 자료는 어떻게 표현할까?

가 모듈 개관

성취기준	12인수02-01 수와 수학 기호를 이용하여 실생활의 텍스트 자료를 목적에 알맞게 표현할 수 있다. 12인수02-02 수와 수학 기호로 표현된 텍스트 자료를 처리하는 수학 원리를 이해하고 자료를 시각화할 수 있다.
모듈 설계 의도	성취기준의 의미 텍스트 자료를 컴퓨터가 처리할 수 있도록 벡터로 표현해 본다. 또한 텍스트 자료를 처리하여 시각화하는 활동을 통해 인공지능에서 텍스트 자료 처리의 원리를 이해할 수 있게 한다. 학습 활동 계획 이 모듈에서는 벡터의 개념을 학습하고 텍스트 자료를 벡터로 바꿔 보는 방법을 학습하며, TD-IDF와 같은 다양한 방법으로 자료를 표현하는 방법의 특징에 대해 학습하고, 텍스트 자료의 빈도를 기반으로 표나 막대그래프, 단어 구름 등으로 시각화하는 것을 목표로 한다. 기대하는 학습 결과 개인별 학습이나 모듈별 탐구 학습을 통해 문서를 벡터로 표현하는 방법을 습득하게 될 것이다. 더 나아가 자료를 다양한 방법으로 시각화하는 방법을 이해하고 텍스트 자료를 처리, 표현하는 방법을 설명할 수 있을 것이다. 모듈 자료에 대한 유의점 공학적 도구를 활용하므로 컴퓨터나 태블릿 PC 등이 사용 가능한 환경에서 수업을 진행해야 한다.
교과 역량	의사소통, 정보 처리, 태도 및 실천
학습 요소	벡터
교수·학습 방법	탐구 학습, 매체 및 도구 활용 학습
교수·학습 방법 및 유의 사항	• 실생활의 텍스트 자료로 댓글이나 감상평 등을 다룰 수 있다. • 자료는 정형, 비정형, 범주형, 연속형 등 여러 가지 유형을 다룰 수 있다. • 텍스트 자료를 수와 수학 기호로 표현할 때 공학적 도구를 이용할 수 있다. • 실생활의 텍스트 자료를 수와 수학 기호를 이용하여 다양한 방식으로 표현해 보고 토론하는 활동을 통해 각 표현 방식의 장단점을 이해하게 한다.

		• 텍스트 자료는 벡터로 표현할 수 있음을 간략하게 소개할 수 있다. • 텍스트 자료를 변형하는 활동을 통해 자료 처리의 원리를 이해하게 하고, 이때 공학 적 도구를 이용할 수 있다. • 텍스트 자료를 표, 그림, 그래프, 단어 구름 만들기 등을 이용하여 시각화한 후 이를 해석해 보는 활동을 할 수 있으며, 이때 공학적 도구를 이용할 수 있다.	
평가 방법		포트폴리오 평가	
평가 방법 및 유의 사항		• 자료의 표현에 대한 평가는 실제 자료를 목적에 맞게 수와 수학 기호로 표현하고 처리 할 수 있는지를 어렵지 않은 수준에서 간단히 다룬다.	
차시	주제	교수·학습 활동	평가 활동
1	텍스트 자료의 표현	탐구 학습 텍스트 자료의 표현 • 자료의 유형, 벡터의 개념 이해하기 • 자료를 벡터로 표현하기 • 단어를 수로 표현하는 방법 이해하기 • 실제 텍스트 자료를 여러 가지 수로 표현 해 보기	포트폴리오 평가 • 벡터 및 로그의 개념을 잘 이해하였는가? • 벡터나 로그를 이용하여 자료를 수나 기호 로 잘 표현하였는가? • 자료를 수와 기호로 표현하는 방법에 대해 이해하였는가?
2	텍스트 자료의 시각화	탐구 학습 텍스트 자료의 시각화 • 텍스트 자료를 시각화하기 전에 전처리 하는 이유에 대해 이해하기 • 텍스트 자료를 시각화하는 다양한 방법 에 대해 학습하기 • 단어 구름으로 텍스트 자료를 시각화 하고 분석해 보기	포트폴리오 평가 • 텍스트 자료를 시각화하는 방법을 이해 하고 단어 구름으로 잘 표현하였는가? • 단어 구름으로 텍스트 자료를 나타내는 원리를 이해하였는가? • 단어 구름으로 나타난 자료를 보고 키워드 분석을 할 수 있는가?
3	텍스트 자료의 시각화 실습	매체 및 도구 활용 학습 텍스트 자료의 시각화 실습 • 텍스트 자료를 처리하는 방법과 원리에 대해 이해하기 • 텍스트 자료를 표, 그래프, 단어 구름 등으로 표현하기 • 텍스트 자료를 시각화한 후 자료 분석 하기	포트폴리오 평가 • 텍스트 자료를 시각화하는 방법과 원리에 대해 알고 있는가? • 표, 그래프, 단어 구름 등으로 표현할 수 있는가?

나 차시별 교수·학습 자료

1 차시 | 텍스트 자료의 표현

학습 목표	벡터와 로그의 의미를 이해한다. 벡터를 이용하여 텍스트 자료를 표현할 수 있으며, 자료를 표현하는 다양한 방법의 의미를 설명할 수 있다.
수업의 의도	수업 의도 실제 접할 수 있는 텍스트 자료를 수와 벡터로 표현해 보고 계산해 보는 활동을 통해 표현 방법을 이해함에 목적이 있다. 이를 위해, 벡터나 로그의 개념을 학습하고, 비정형 자료를 어떻게 컴퓨터가 인식하는지를 설명할 수 있게 한다. 수업의 주안점 학생들이 자료를 표현하는 방법에 대해 올바르게 이해하고, 제시된 자료를 벡터나 수로 올바르게 표현하고, 그렇게 표현하는 이유를 토론 활동 등을 통해 설명할 수 있도록 한다. 또한, 계산 시에는 공학 도구를 이용할 수 있도록 하며 계산 과정의 의미에 대하여 의사소통할 수 있게 한다. 수업 활동 시 유의점 TF-IDF의 값, log의 값은 계산기나 컴퓨터를 도입하여 계산하되 활동이 계산 중심이 되지 않게 한다. 평가의 주안점 단순히 자료를 DF, IDF와 같은 여러 가지 값을 계산하는 데 그치지 않고, 그 값을 사용하는 이유에 대해 올바르게 이해하는 데 중점을 두어 평가한다.
준비물	활동지, 컴퓨터

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	3인 1조로 모둠 구성 - 활동지, 자기 평가 및 동료 평가지 배부	
수행	모둠 활동	- 자료의 유형, 벡터의 의미 이해하기 - 텍스트 자료를 벡터로 표현하기 - 로그의 의미 이해하기 - 텍스트 자료에 대하여 DF, IDF, TF-IDF의 의미 이해하기	포트폴리오 평가
정리		자기 평가 및 동료 평가지 작성	



텍스트 자료의 표현

년

월

일



자료의 종류에 대해 알아보자.

- 1) 정형 자료(Structured Data): 정형 자료는 정해진 규칙에 맞게 자료를 수집하여 수치나 이름, 종류 등 간단한 형식으로 나타낸 것이다. 자료의 값을 보고 그 의미를 파악하기 쉽게 일반적으로 아래와 같이 표로 정리하여 나타낸다.

이름	성별	나이	거주지	직업	스마트폰 이용 요금(단위: 천 원)	스마트폰 자료 사용량(G)
최민준	M	32	수원	회사원	90	8
방정화	F	18	서울	학생	47	2
김하늘	M	28	마포	자영업	55	3
...	...	...	...	...	...	...

자료와 자료 사이에 끊어짐이 없는 값을 취하는 자료를 연속형 자료(Continuous Data)라고 하고, 자료 사이에 끊어짐이 있는 자료를 이산형 자료(Discrete Data)라고 한다. 또, 자료들 사이에 순위가 있는 경우는 순위형 자료(Ordinal Data)라고 하며, 자료 간 순위가 없는 경우는 명목형 자료(Nominal Data)라고 한다.

- 2) 비정형 자료(Unstructured Data): 비정형 자료는 정형 자료와 달리 일정한 규칙이 없어서 정형 자료처럼 정리하지 않고 원래 있는 그대로 수집한 자료이다. 예시로는 영화 리뷰, 상품 평가 댓글, 동영상과 같은 예가 있다.

좋아요 갤럭시탭!

옵션 없음

테블릿 PC 처음 쓰는데 화질도 좋고 필기감도 좋고 너무 좋아요!

추천합니다

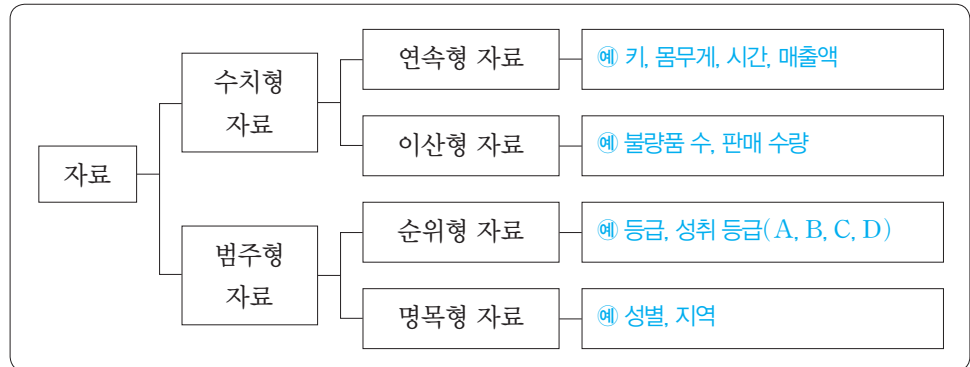
옵션 없음

연후로 택배 대란이라는 뉴스를 봐서 걱정했는데 빨리 왔습니다. 택배 파손된 부분 없이 안전하게 도착했고, 상품도 잘 포함되어 왔습니다. 이것저것 테스트 해 봤는데 다 정상이고 좋습니다. 필기감도 좋고,

[상품 평가 댓글]

1 정형 자료와 비정형 자료의 예를 찾아보자.

1) 정형 자료



2) 비정형 자료

예 텍스트 자료(영화, 상품 리뷰), 음성 자료(녹음 파일), 영상 자료(영화)



벡터에 대해 알아보자.

이름	성별	나이	거주지	직업	스마트폰 이용 요금(단위: 천 원)	스마트폰 자료 사용량(G)
최민준	M	32	수원	회사원	90	8
방정화	F	18	서울	학생	47	2
김하늘	M	28	마포	자영업	55	3
...	...	...	...	...	...	...

위의 표에서 세 사람의 나이, 요금, 자료 사용량을 순서쌍으로 $(32, 90, 8)$, $(18, 47, 2)$, $(28, 55, 3)$ 과 같이 나타낼 수 있다. 이처럼 자료 여러 개를 순서쌍의 형태로 표현한 것을 벡터라고 한다. 이때 나이, 요금, 자료 사용량을 벡터를 구성하는 성분이라고 한다. 성분이 a_1, a_2 의 2개로 이루어진 벡터는 (a_1, a_2) , 성분이 a_1, a_2, a_3 의 3개로 이루어진 벡터는 (a_1, a_2, a_3) 과 같이 나타낸다. 마찬가지로 $a_1, a_2, \dots, a_n$ 의 n 개의 성분으로 이루어진 벡터는 $(a_1, a_2, \dots, a_n)$ 과 같이 나타낸다.

또한 벡터의 덧셈과 뺄셈은 각각의 성분의 합과 차로 정의하고, 벡터의 실수배는 각각의 성분에 실수배를 한 것으로 정의한다.

예를 들어, 두 벡터 $(a_1, a_2), (b_1, b_2)$ 에서

$$\begin{aligned} & \bullet (a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 + b_2), (a_1, a_2) - (b_1, b_2) = (a_1 - b_1, a_2 - b_2) \\ & \bullet k(a_1, a_2) = (ka_1, ka_2) \text{ (단, } k \text{는 실수)} \end{aligned}$$

 텍스트 자료를 단어사전(백오브워즈, Bag of Words, BoW)을 이용하여 벡터로 표현해 보자.

단어사전이란 텍스트 문서에 있는 단어들을 모은 집합이다. 즉, 단어들의 출현 빈도에만 집중하는 텍스트 자료의 표현 방법이다.

문서 1: ‘수학은 아름답다.’

문서 2: ‘수학은 중요하다. 인공지능에서도 수학은 중요하다.’

위의 문서 1의 단어사전을 나타내는 집합은

$A = \{\text{수학은, 아름답다}\}$ 이고,

문서 2의 단어사전을 나타내는 집합은

$B = \{\text{수학은, 중요하다, 인공지능에서도}\}$ 이다.

문서 1과 문서 2를 모두 포함하는 단어사전은

$A \cup B = \{\text{수학은, 중요하다, 아름답다, 인공지능에서도}\}$ 라고 할 수 있다.

이때, 문서를 단어사전에 나타낸 단어의 빈도를 성분으로 하는 벡터로 표현할 수 있다.

예를 들면, 문서 1, 문서 2를 벡터로 표현하면 각각 (1, 0, 1, 0), (2, 2, 0, 1)과 같이 표현할 수 있으며 이를 단어사전 벡터라고 한다. 이때, 벡터의 성분은 단어사전의 단어 순서와 같다.

1 영화평, 상품평, 기사 리뷰 등에서 관련 문장 2개를 찾아보고, 각각의 문장에 대하여 두 문장에 포함된 모든 단어로 이루어진 단어사전을 만들고, 벡터로 표현해 보자.

예	
문서 1	명량, 역사를 바꾼 가장 위대한 이순신 장군
문서 2	두려움을 용기로 바꾼 이순신
단어사전	{가장, 두려움을, 명량, 바꾼, 역사를, 위대한, 용기로, 이순신, 장군}
문서 1의 벡터	(1, 0, 1, 1, 1, 1, 0, 1, 1)
문서 2의 벡터	(0, 1, 0, 1, 0, 0, 1, 1, 0)

2 단어사전을 만들고 벡터로 표현하는 방법의 장단점에 대해 생각해 보자.

1) 다음 대화를 보고 단어사전을 만들고 벡터로 표현하는 방법의 특징에 대하여 이야기해 보자.



'나는 엄마를 사랑해'일 수도 있지만, '엄마는 나를 사랑해'일 수도 있다.

단어사전을 이용하여 벡터를 만든 경우 단어의 출현과 빈도로 문장의 의미를 대략적으로 추론할 수는 있지만, 단어의 순서는 알 수 없기 때문에 원문을 알 수 없다.

즉, 다르게 해석할 여지가 있다.

2) 단어사전을 이용하는 방법의 특징에 대해서 이야기해 보자.

- 문장을 어렵지 않게 벡터로 표현할 수 있다.
- 실제로 굉장히 많은 단어를 나타내기에 너무 힘들다. (차원이 너무 커진다.)
- 1)의 예시와 같이 다른 의미를 가진 문장을 같은 벡터로 표현하는 경우가 생긴다.
- 새로운 단어가 나타나는 경우에는 다시 벡터를 조정해야 한다. (차원을 늘려야 한다.)
- '지우는 모든 사람들에게 사랑을 받는다.'라는 문장과 '모든 사람은 지우를 사랑한다.'라는 문장의 벡터 표현은 다르지만 동일한 의미를 지니는데 이는 구분하지 못한다.

log에 대해 알아보자.

예전의 천문학자는 별자리 관측과 같은 자료를 바탕으로 계산하는 경우 큰 수의 복잡한 계산을 비교적 간단하게 하기 위해 로그(log)라는 개념을 사용하였다.

$a > 0, a \neq 1$ 일 때, 양수 N 에 대하여

$$a^x = N$$

을 만족시키는 실수 x 는 오직 하나 존재한다. 이 실수 x 를

$$x = \log_a N$$

과 같이 나타내고, a 를 밑으로 하는 N 의 로그라고 한다.

예를 들어,

$2^x = 4$ 를 만족시키는 값은 $x = \log_2 4 = 2$, $2^x = 5$ 를 만족시키는 값은 $x = \log_2 5$,

$2^x = 8$ 을 만족시키는 값은 $x = \log_2 8 = 3$ 이다.

특히, 밑을 10으로 하는 로그를 상용로그라고 하며 이 경우는 밑을 생략하여

$\log_{10} N = \log N$ 으로 나타낸다.

위에서 알 수 있듯이 어떤 수에 log를 취하면 그 값은 작아지므로 큰 수의 계산도 비교적 간단하게 할 수 있음을 알 수 있다. 또한 크기가 다른 어떤 두 수에 log를 취하면 큰 수의 log의 값이 더 크다.

1 다음 ☐ 안에 알맞은 수를 써 보자.

1) $3^x = 7$ 일 때, $x =$

2) $5^x = 125$ 일 때, $x =$

3) $\log_3 81 =$

4) $\log 100 =$

1) $x = \log_3 7$

2) $x = 3$

3) $3^4 = 81$ 이므로 로그의 정의에 따라 $\log_3 81 = 4$ 이다.

4) $10^2 = 100$ 이므로 로그의 정의에 따라 $\log 100 = 2$ 이다.



여러 개의 문서를 수와 기호로 표현하는 방법을 알아보자.

하나의 문서보다는 여러 문서를 통해 중요한 정보를 찾을 필요가 있다. 여러 개의 텍스트 문서에서 중요한 단어나 주제어로 적합한 단어를 찾는 방법에 대해 생각해 보자.

다음은 영화 ‘정말 최고의 영화’, ‘너무 너무 재밌는 영화’의 감상평 텍스트 자료를 행에는 사람들의 영화 감상평을, 열에는 출현한 단어들의 빈도수를 표현한 것이다.

	정말	너무	최고의	재밌는	영화
감상평 1	1	0	1	0	1
감상평 2	0	2	0	1	1

1) TF(Term Frequency): 각 문서에서 용어가 나타난 빈도이다.

예 위의 표에서 ‘너무’라는 단어의 감상평 2에서 TF의 값은 2이다.

2) DF(Document Frequency): 어떤 단어가 몇 개의 문서에서 나타났는지에 대한 값이다.

예 위의 표에서 ‘정말’이라는 단어의 DF의 값은 1이고, ‘영화’라는 단어의 DF의 값은 2이다.

3) IDF(Inverse Document Frequency): 역문서 빈도라고 하며,

$IDF = \log\left(\frac{n}{DF}\right)$ 이다. (단, n 은 문서의 개수)

예 ‘정말’이라는 단어의 DF의 값은 1이므로 IDF의 값은 $\log\left(\frac{2}{1}\right) = \log 2$ 이다.

4) TF-IDF(Term Frequency-Inverse Document Frequency): 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 값이며,

$TF-IDF = TF \times \log\left(\frac{n}{DF}\right)$ 이다.

예 ‘정말’이라는 단어의 감상평 1에서 TF의 값은 1이고, DF의 값은 1이므로 TF-IDF의 값은 $1 \times \log\left(\frac{2}{1}\right) = \log 2$ 이다.

1 다음 기사의 일부를 보고 답해 보자.

“오늘부터 마스크 꼭 쓰세요”... 위반하면 최대 10만 원 과태료

- ① 이제는 대중교통이나 의료 기관, 약국 등 사람들이 자주 찾는 시설에서 마스크를 쓰지 않았다가 적발되면 과태료를 내야 합니다.
- ② 현행 ‘사회적 거리두기’ 1단계에서 마스크를 반드시 써야 하는 시설 및 장소는 중점·일반 관리 시설 23종과 대중교통, 집회·시위장, 의료 기관·약국, 요양 시설 및 주야간 보호 시설, 종교 시설 등입니다.
- ③ 이 가운데 중점 관리 시설 9종에는 클럽-룸살롱 등 유흥 시설 5종·노래 연습장·실내 스탠딩 공연장·방문 판매 등 직접 판매 홍보관·식당-카페 등이, 일반 관리 시설 14종에는 PC방·학원(교습소 포함)·독서실·스터디 카페·결혼식장·장례식장·영화관·공연장·목욕장업·직업 훈련 기관 등이 포함됩니다.
- ④ 의약외품으로 허가된 마스크가 없으면 입과 코를 완전히 가릴 수 있는 천 마스크나 면 마스크, 일회용 마스크를 써도 되지만 망사형 마스크, 밸브형 마스크는 인정되지 않습니다.
스카프 등의 옷가지로 얼굴을 가리는 행위도 마스크 착용으로 인정되지 않아 과태료를 물 수 있습니다.
- ⑤ 마스크를 썼더라도 ‘턱스크’, ‘코스크’ 등 입이나 코를 완전히 가리지 않았다면 마스크를 쓰지 않은 것으로 간주돼 과태료가 부과될 수 있습니다. 위반 당사자에게는 횡수와 관계없이 최대 10만 원의 과태료가 부과됩니다.

[출처: https://news.sbs.co.kr/news/endPage.do?news_id=N1006071972&plink=SEARCH&cooper=SBSNEWSSEARCH&plink=COPYPASTE&cooper=SBSNEWSSEND]

1) 위 기사에 쓰인 각각의 문장에 대하여 TF의 값을 찾아 표를 완성해 보자.

	마스크	시설	학원	망사형	턱스크	과태료	카페	10만 원
문장 ①	1	1	0	0	0	1	0	0
문장 ②	1	5	0	0	0	0	0	0
문장 ③	0	3	1	0	0	0	2	0
문장 ④	7	0	0	1	0	1	0	0
문장 ⑤	2	0	0	0	1	2	0	1

1-1) 문장별 단어사전 벡터를 만들어 보자. 또, 문장 전체의 단어사전 벡터를 벡터의 합을 이용하여 구해 보자.

문장 ①의 벡터: (1, 1, 0, 0, 0, 1, 0, 0)

문장 ②의 벡터: (1, 5, 0, 0, 0, 0, 0, 0)

문장 ③의 벡터: (0, 3, 1, 0, 0, 0, 2, 0)

문장 ④의 벡터: (7, 0, 0, 1, 0, 1, 0, 0)

문장 ⑤의 벡터: (2, 0, 0, 0, 1, 2, 0, 1)

문장 ①~⑤의 벡터는 모든 벡터의 합과 같으므로 (11, 9, 1, 1, 1, 4, 2, 1)이다.

1-2) 가장 빈도수가 높은 단어와 주제어로 적합한 단어라고 생각하는 단어를 각각 말해 보자.

• 가장 빈도수가 높은 단어: 마스크

• 주제어로 적합한 단어: ☞ 과태료

마스크가 가장 빈도수가 높지만 실질적으로 중요한 단어라고 할 수 없다.

상황에 따라 달라지겠지만, 빈도수가 높다고 무조건 중요한 단어라고 할 수 없다.

빈도수가 4인 과태료는 빈도수도 높은 편이며 주제어로 적합하다고 생각한다.

2) DF, IDF의 값을 구해 보자.

	마스크	시설	학원	망사형	텍스크	과태료	카페	10만 원
DF	4	3	1	1	1	3	1	1
IDF	0.097	0.222	0.699	0.699	0.699	0.222	0.699	0.699

2-1) DF의 값이 크다고 중요한 단어라고 할 수 있는지 이야기해 보자.

DF의 값이 크다고 반드시 중요한 단어라고 할 수는 없다. 마스크와 같이 문서에 자주 등장하는 말이 반대로 유의미한 단어라고 말할 수 없는 경우도 있다. 오히려 특정한 문장에만 나타나는 단어가 중요한 의미를 지닐 수 있다.

2-2) IDF에서 $\frac{n}{DF}$ 의 값을 구하지 않고, $\log\left(\frac{n}{DF}\right)$ 의 값으로 나타내는 이유에 대해 이야기해 보자.

실제로 자료를 분석할 경우에는 많은 양의 문서를 분석하게 된다. 즉, n 의 값이 커질 것이다.

반대로 문서에 중복하여 등장하는 단어의 빈도수는 그에 비례하게 높지 않아 $\frac{n}{DF}$ 의 값이 상대적으로 큰 수로 증가할 것이다. 이때, $\log$ 를 취하면 그 값을 적당한 값으로 조정할 수 있다.

3) TF-IDF의 값을 구해 보자.

	마스크	시설	학원	망사형	텍스크	과태료	카페	10만 원
문장 ①	0.097	0.222	0	0	0	0.222	0	0
문장 ②	0.097	1.11	0	0	0	0	0	0
문장 ③	0	0.666	0.699	0	0	0	1.398	0
문장 ④	0.679	0	0	0.699	0	0.222	0	0
문장 ⑤	0.194	0	0	0	0.699	0.444	0	0.699

3-1) TF-IDF의 값을 참고하여 문장에서 단어의 중요성에 대하여 이야기해 보자.

문장 ④에서 마스크의 TF-IDF의 값이 크긴 했지만 나머지 문서에서의 값은 작다. 즉, 모든 문서에 공통적으로 들어있는 마스크는 중요성이 떨어진다고 할 수 있다.

문장 ①에서는 시설이나 과태료, 문장 ②에서는 시설, 문장 ③에서는 학원, 카페, 문장 ④에서는 망사형, 문장 ⑤에서는 텍스크, 10만 원 등이 오히려 핵심 단어라고 추론할 수 있다.

모든 문서에 등장하는 단어는 대체로 중요도가 높다고 할 수 없다. 즉, 특정한 문장에만 나타나는 단어가 중요하다고 할 수 있으며 특정한 문장에만 자주 등장하는 단어가 좀 더 중요하다고 할 수 있다.



2 AI 웹실험실 을 이용하여 TF-IDF의 값을 구해 보자.

TF-IDF(Term Frequency-Inverse Document Frequency) 계산

index	공감	꼭	남녀	당시	더	만	보고	사람	사랑	생각	수년	영화	인구	자체	장면	재생산	전개	주인공	타이타닉	회상뿐
문장1	0	0	0	0	3.22	1.61	0	0	0	0	0	0.71	0	0	0	0	0	0	1.2	0
문장2	0	1.61	0	0	0	0	3.22	0	0	2.41	1.61	0	0	0	0	0	0	0	0	0
문장3	0	0	0	0.92	0	0	0	1.2	0	0	0	0.36	0	0	0	0	0	0	0	0
문장4	0	0	0	0.92	0	0	0	0	0	0	0	0.71	1.61	1.2	3.22	1.61	0	0	1.2	0
문장5	0	0	0	0	0	0	0	0	0	0	0	0.36	0	0	0	0	0	0	0	0
문장6	0	0	0.92	0.92	0	0	0	2.41	1.2	0	0	0.36	0	0	0	0	0	1.2	0	1.2
문장7	1.2	0	0.92	0	0	0	0	0	1.2	0	0	0	0	0	0	0	1.61	1.2	0	0
문장8	1.2	0	0	0	0	0	0	0	0	0	0	0.36	0	0	0	0	0	0	0	1.2
문장9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
문장10	0	0	0.92	0	0	0	0	0	0	2.41	0	0	0	1.2	0	0	0	0	0	0

[웹실험실의 TF-IDF 계산 예시]

지도의 주안점

- 교사는 자료를 수나 기호로 나타내는 의미를 토론하는 활동을 통해 이해하도록 도움을 준다. 또한 계산의 경우 컴퓨터 프로그램 등을 적극적으로 활용하도록 한다.

핵심 발문

- 텍스트 문서를 벡터로 나타내는 방법의 장단점은 무엇일까?
- TF-IDF의 의미와 계산 방법은 무엇일까?
- TF-IDF를 계산할 때, 계산기 혹은 컴퓨터를 어떻게 이용할 수 있을까?

교과서 연계 방안

- 교과서를 통해 텍스트 자료를 벡터로 표현하는 방법을 충분히 이해한 후 활동할 수 있도록 한다.



관련 교과 역량

- 의사소통
 - 텍스트 자료를 수 또는 벡터로 표현하는 방법에 대해 이야기할 수 있는가?
- 태도 및 실천
 - 자료를 수 또는 벡터로 표현하는 것에 어려움을 겪고 있는 모둠원을 배려하고 존중할 수 있는가?

평가 결과 피드백

- 텍스트 자료를 수나 기호로 표현하고, 다양한 의미의 빈도수를 계산하는 방법의 의미를 정확하게 이해하고 있는지 확인하기 위해 설명하거나 토론하는 활동을 하는 기회를 제공한다.
- 수나 기호로 표현한 자료를 분석하기 위해 시각화하는 과정에서 어떤 수학적 원리를 이용했는지 이해하고 있는지 확인한다.

2차시 | 텍스트 자료의 시각화

학습 목표	자료를 시각화할 수 있는 다양한 표현 방법을 학습하고 단어 빈도수의 벡터로부터 단어 구름을 시각화하는 과정을 설명할 수 있다.
수업의 의도	수업 의도 텍스트 자료를 시각화하는 다양한 방법을 알고, 단어 빈도수의 벡터로부터 단어 구름으로 시각화하는 방법을 이해한다. 수업의 주안점 텍스트 자료의 시각화를 위해 텍스트 자료를 표현하는 방법을 먼저 학습한다. 또한 단어 빈도수의 벡터로부터 단어 구름을 그려 주제어를 찾는 시각화 과정을 이해하고 텍스트 자료를 분석할 수 있다. 수업 활동 시 유의점 단순히 자료를 시각화하는 활동에 중점을 두기보다는 시각화하여 표현하는 원리에 대해 이해하도록 한다. 평가의 주안점 단어 구름을 나타내는 원리를 이해하고 텍스트 자료를 분석할 수 있는지를 평가한다.
준비물	활동지, 노트북(태블릿 PC나 스마트폰으로 대체 가능)

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	2~3명을 1개 모둠으로 구성	
수행	개별 활동	- 다양한 시각화 방법 탐구하기 - 시각화 방법의 유형 탐구하기	포트폴리오 평가
	모둠 활동	- 단어 빈도수의 벡터로부터 단어 구름 시각화 방법을 이해하고 단어 구름 만들기 - 단어 구름의 장점과 특징에 대한 이해를 바탕으로 자료 분석하기	
정리		단어 구름 시각화 방법에 대해 정리	



텍스트 자료의 시각화

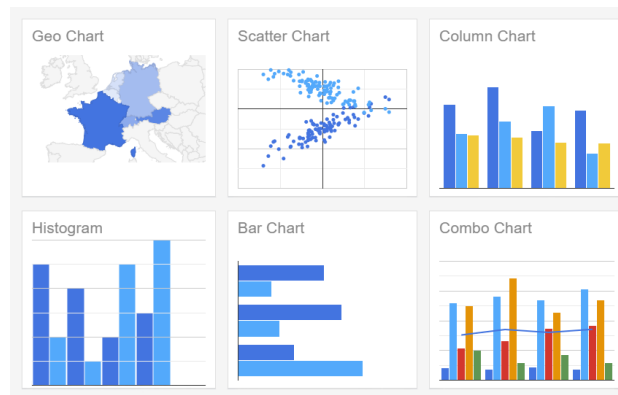
년 월 일



다양한 시각화 방법에 대해 알아보자.

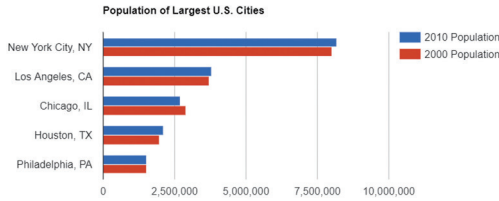
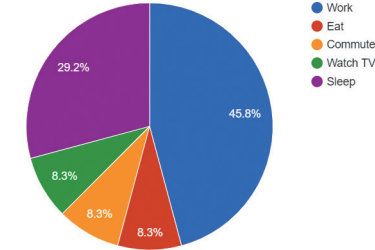
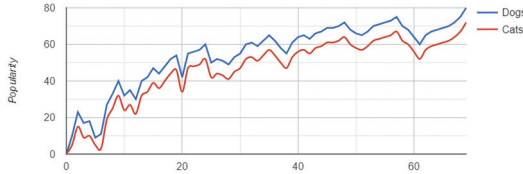
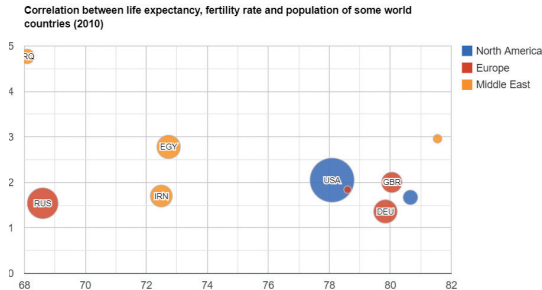
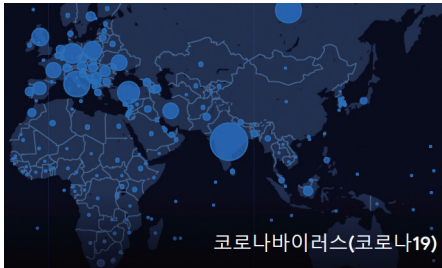


[출처: <https://datavizproject.com/>]



[출처: <https://developers.google.com/chart>]

1 다음과 같은 시각화 방법은 어떤 장점이 있는지 생각해 보고, 예시를 찾아보자.

시각화 방법	장점과 예시
 bar chart	• 장점: 비교하고 순위를 매기는 데 효율적이다. • 예시: 남녀 공무원 임용 현황 등
 pie chart	• 장점: 전체에서 차지하는 비중을 한 눈에 알 수 있다. • 예시: 우리나라 회사의 세계 시장 점유율 등
 Line chart	• 장점: 트렌드와 변화의 추세를 알 수 있다. • 예시: 연도별 집값의 변화 추이 등
 Bubble chart	• 장점: 상관관계를 파악하기 쉽다. • 예시: 상품 간의 상관관계, 나이와 마라톤 기록의 관계 등
 Geo chart	• 장점: 자료의 흩어진 정도를 쉽게 알 수 있다. • 예시: 지도를 보고 구매자가 어디에 많은지, 어느 지역에 가게가 많은지 등

[출처: <https://developers.google.com/chart>]



워드 클라우드(단어 구름)를 이용해서 텍스트를 분석해 보자.

워드 클라우드(단어 구름)는 빈도가 높은 단어는 크게, 그렇지 않은 단어는 작게 보이도록 나타내는 시각화 방법 중의 하나이다. 즉, 여러 문서 또는 하나의 문서를 단어사전 벡터로 나타내고 성분의 값이 큰 글자는 크게, 성분의 값이 작은 글자는 작게 나타낸 것이다. 워드 클라우드 사이트(<http://wordcloud.kr>)에서 만들어 볼 수 있다.



1 단어사전 벡터를 이용하여 워드 클라우드를 만들 때 글자의 폰트를 정하여 보자. 아래 단어사전 벡터를 참고하여 답해 보자.

단어사전	{증상, 마스크, 사람, 이상, 기침, 손을, 수칙, 실내}
문장 ①의 벡터	(2, 2, 3, 2, 0, 1, 0, 0)
문장 ②의 벡터	(4, 1, 2, 1, 1, 1, 0, 1)
문장 ③의 벡터	(3, 1, 2, 2, 1, 1, 2, 0)

1) 문장 ①~③을 합한 전체의 단어사전 벡터를 구해 보자.

각각의 문장의 벡터에서 성분의 합을 구하면 (9, 4, 7, 5, 2, 3, 2, 1)이다.

2) 워드 클라우드로 변환할 때, 글자별로 폰트를 어떻게 정하면 좋을지 이야기해 보자.

전체 문장의 단어사전 벡터에서 성분의 수가 큰 증상을 가장 큰 폰트로 정하고, 사람, 이상, 마스크 순으로 비율을 계산하여 순차적으로 작게 만들며, 실내를 가장 작은 폰트로 정한다.

2 위의 워드 클라우드는 어떤 자료를 넣어서 얻은 자료인지 추측해 보고, 워드 클라우드의 장점에 대해 이야기해 보자.

- 해석: 증상, 마스크, 기침 등의 단어가 자주 등장하는 자료임을 알 수 있어, 코로나19 방역 지침과 관련된 자료라고 추측할 수 있다.
- 장점: 문서에 자주 등장하는 키워드를 중심으로 시각화한 자료이므로 문서를 읽지 않아도 내용을 한 눈에 예측할 수 있다.

생활동행 워드 클라우드에 넣은 자료는 생활 속 거리두기 지침 중 개인 방역 5대 중요 수칙이다.

3 워드 클라우드에 접속하여 뉴스 기사 등의 자료를 입력하고 구름 단어 모양을 만들어 보자.
그리고 친구와 서로 바꾸어 어떤 내용의 기사인지 추측해 보자.



해석: 빈도수가 높은 단어는 코로나19, 백신, 생산, 계약, 치료제, 위탁 생산, 개발, 임상 등이며 이를 근거로 추측해 보면 코로나19의 임상 시험을 거친 백신, 치료제 개발과 관련한 글이며 개발, 생산, 계약 등의 단어로 이루어 보아 실제 백신이나 치료제를 만들 것이라는 글로 추측해 볼 수 있다.

• **활동팁** 위의 워드 클라우드에 넣은 기사의 제목은 “백신보다 한발 앞선 코로나 치료제… 생산 기지 주목 받는 한국”이며 기사 출처는 <https://news.joins.com/article/239227780>이다.

4 1의 워드 클라우드와 3에서 자신이 만든 워드 클라우드를 참고하여 답해 보자.



[1의 워드 클라우드]



[자신이 만든 워드 클라우드]

1) 1의 워드 클라우드와 자신이 만든 워드 클라우드에 나타난 단어 중에서 의미가 적은 단어를 각각 적어 보자.

- 1의 워드 클라우드: **합니다, 또는, 있습니다** 등
- 자신이 만든 워드 클라우드: **것으로, 있다** 등

2) 워드 클라우드 같은 자료를 만들 때 주의해야 할 부분에 대해서 생각해 보고 이야기해 보자.

- ‘나는’, ‘있다’ 등과 같이 어느 문서에나 자주 등장하는 단어는 큰 의미가 없는 경우가 많으므로 제거하는 것이 좋다.
- ‘사람과’, ‘증상’과 같은 단어를 보여 주기보다는 조사를 제거하고 ‘사람’, ‘증상’ 등으로 제시하는 것이 좋을 것이다. 만약 ‘증상’이, ‘증상’과 같은 단어가 나온다면 한 가지 경우로 보는 게 더욱 타당하다고 생각한다.

지도의 주안점

- 교사는 자료를 수나 기호로 나타내는 의미를 토론하는 활동을 통해 이해하도록 도움을 준다. 또한 계산의 경우 컴퓨터 프로그램 등을 적극적으로 활용하도록 한다.

핵심 발문

- 시각화 방법은 어떤 종류가 있을까?
- 단어 구름으로 시각화하는 방법의 기본 원리는 무엇일까?
- 다양한 시각화 프로그램을 어떻게 이용할 수 있을까?

교과서 연계 방안

- 교과서를 통해 텍스트 자료를 벡터로 표현하는 방법을 충분히 이해한 후 활동할 수 있도록 한다.



관련 교과 역량

- 의사소통
 - 텍스트 자료를 대수적으로 표현하는 방법을 이해할 수 있는가?
- 태도 및 실천
 - 자료를 수 또는 벡터로 표현하는 것에 어려움을 겪고 있는 모둠원을 배려하고 존중할 수 있는가?

평가 결과 피드백

- 텍스트 자료를 수나 기호로 표현하고, 다양한 의미의 빈도수를 계산하는 방법의 의미를 정확하게 이해하고 있는지 확인하기 위해 설명하거나 토론하는 활동을 하는 기회를 제공한다.
- 수나 기호로 표현한 자료를 분석하기 위해 시각화하는 과정에서 어떤 수학적 원리를 이용했는지 이해하고 있는지 확인한다.

3차시 | 텍스트 자료의 시각화 실습

학습 목표	텍스트 자료의 시각화를 위해 텍스트 자료를 원그래프, 막대그래프 등 다양한 그래프로 나타내 본다.
수업의 의도	수업 의도 텍스트 자료를 시각화하는 원리를 이해하고 원그래프, 막대그래프 등 다양한 표현을 이용하여 나타내 본다. 수업의 주안점 목적에 맞는 시각화 방법을 선택하고 텍스트 자료를 시각화할 수 있도록 한다. 시각화된 자료를 토대로 텍스트 자료를 분석해 본다. 수업 활동 시 유의점 시각화 활동을 통해 시각화의 유용성이나 목적과 이유를 생각할 수 있도록 한다. 평가의 주안점 상황에 맞게 그래프로 표현하였는지, 텍스트 자료를 잘 표현하고 분석하였는지 평가한다.
준비물	활동지, 노트북(태블릿 PC나 스마트폰으로 대체 가능)

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	2~3명을 1개 모둠으로 구성	
수행	개별 활동	- 텍스트 전처리 이유와 방법에 대해 알아보기 - 텍스트 분석하기 전의 절차에 대해 알아보기	포트폴리오 평가
	모둠 활동	- 다양한 시각화 방법으로 표현해 보기 - 구름 단어의 장점과 특징에 대한 이해를 바탕으로 자료를 분석하고 발표해 보기	
정리		문서를 다양한 방법으로 시각화하는 원리에 대해 정리	



텍스트 자료의 시각화 실습

년 월 일



텍스트 자료를 분석하기 위한 단계별 절차에 관하여 알아보자.

텍스트 분석 대상이 되는 코퍼스 준비하기

비정형 텍스트 자료들을 준비한다. 문헌 분석에서 그 문헌들의 집합을 코퍼스(corpus)라 부른다. 경우에 따라 적절한 방법을 사용하여 해당 문헌을 수집하고 정리해야 한다.



코퍼스에 대해 숫자나 문장 부호 등을 제거하기

코퍼스에서 숫자나 문장 부호 등을 제거한다. 숫자나 문장 부호가 나타내는 문장의 뉘앙스가 있겠지만, 텍스트 분석은 다량의 문헌에서 대략적인 의미를 파악하는 것을 목표로 하므로 해석하는 경우의 수를 단순화하고자 숫자나 문장 부호를 제거하는 것이 좋다.



불용어 제거하기

텍스트 분석에서 큰 의미가 없으면서 가장 빈번하게 사용되는 단어가 있다. 영어에서는 the, of, and, to 등이 있고, 한글에서는 조사나 접속사 등이 이에 해당하는데, 이를 불용어(stop words)라 한다. 이러한 불용어는 영어 기준으로 대략 400~500개 정도가 있다.



어간 추출하기

어간 추출(stemming)은 단어의 어간을 찾는 기법이다. 즉, 같은 의미인데 문장에서 여러 형태로 사용되는 단어의 경우에 어간을 찾아서 한 단어로 치환하는 역할을 한다. 예를 들어, user, used, using이라는 단어에는 모두 use라는 의미가 있기 때문에 use라는 단어로 치환한다. 마찬가지로 이런 과정을 거치면 분석하는 시간을 줄일 수 있다.



문서를 나타내는 벡터 생성하기

이렇게 정리된 코퍼스를 대상으로 문헌별 용어가 얼마나 나왔는지를 알려 주는 벡터를 생성한다. 이때 앞서 배운 상황에 따라 다양한 값, 예를 들어, TF기준, TF-IDF기준 등으로 나타낸다.

[출처: 이원상, 『머신 러닝을 위한 수학 with 파이썬, R』]

참고 한국어/영어 불용어의 예시

아	어찌됐든	하기보다는	a	ourselves
휴	그위에	차라리	about	out
아이구	게다가	하는 편이 낫다	above	over
아이쿠	점에서 보아	흐흐	after	own
아이고	비추어 보아	놀라다	again	same
어	고려하면	상대적으로 말하자면	against	shan't
나	하게될것이다	마치	all	she
우리	알것이다	아니라면	am	she'd
저희	비교적	췌	an	she'll
따라	좀	그렇지 않으면	and	she's
의해	보다더	그렇지 않다면	any	should
을	비하면	안 그러면	are	shouldn't
를	시키다	아니었다면	aren't	so
에	하게하다	하든지	as	some
의	할만하다	아니면	at	such
가	의해서	이라면	be	than
으로	연이서	좋아	because	that
로	이어서	알았어	been	that's
에게	잇따라	하는것도		
뿐이다	뒤따라	의거하여		

[출처: <https://www.ranks.nl/stopwords>]



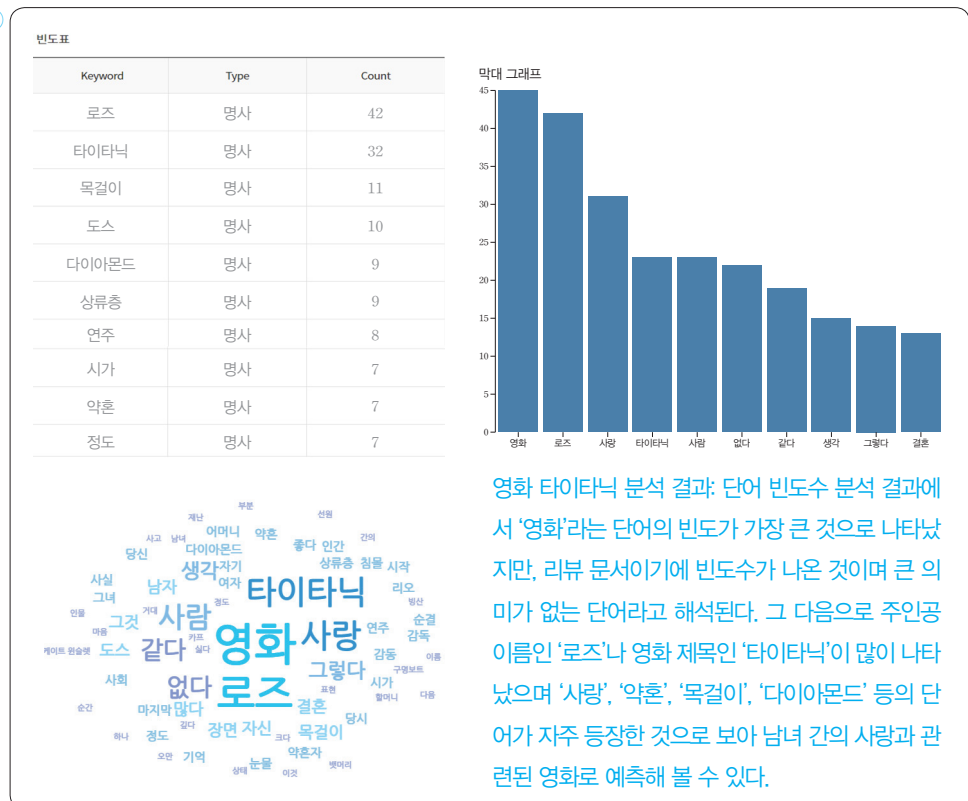
자료에서 필요한 단어를 분류하는 과정에 대한 예시를 알아보자.

자료	리뷰
원본	텍스트 분석을 위해서 프로그램을 이용합니다.
최소 의미 단위 분류(토큰화)	텍스트, 분석, 을, 위해, 서, 프로그램, 을, 이용, 합니다.
불용어 제거	텍스트, 분석, 위해, 프로그램, 이용, 합니다.
명사 추출	텍스트, 분석, 프로그램, 이용

● 통찰 토근화 과정은 최소 의미 단위로 분류하는 과정이며 정제된 형태의 자료로 변형시켜야 벡터의 차원도 줄어듦과 자료를 처리하는 시간도 줄일 수 있음을 이해하도록 한다.



1 **AI 웹실험실**을 이용하여 영화 리뷰, 상품 리뷰, 뉴스 기사 등의 자료를 여러 가지 방법으로 시각화하여 텍스트 자료를 분석해 보고, 그 결과에 대하여 발표해 보자.



영화 타이타닉 분석 결과: 단어 빈도수 분석 결과에서 ‘영화’라는 단어의 빈도가 가장 큰 것으로 나타났지만, 리뷰 문서이기에 빈도수가 나온 것이며 큰 의미가 없는 단어라고 해석된다. 그 다음으로 주인공 이름인 ‘로즈’나 영화 제목인 ‘타이타닉’이 많이 나타났다. ‘사랑’, ‘약혼’, ‘목걸이’, ‘다이아몬드’ 등의 단어가 자주 등장한 것으로 보아 남녀 간의 사랑과 관련된 영화로 예측해 볼 수 있다.

활동Tip 학생들이 보지 않았던 영화 등을 시각화 자료를 통해 분석, 예측해 보는 활동을 통해 직접 책을 읽거나 영화를 보지 않아도 자료 분석만으로도 주요 내용을 예측할 수 있음을 알도록 한다. 실제로 무수히 많은 양의 문서를 다양한 분석 기법을 통해 분류하는 것도 비슷한 맥락이라고 할 수 있다.

지도의 주안점

- 교사는 자료를 수나 기호로 나타내는 의미를 토론하는 활동을 통해 이해하도록 도움을 준다. 또한 계산의 경우 컴퓨터 프로그램 등을 적극적으로 활용하도록 한다.

핵심 발문

- 텍스트 자료의 처리 절차는 무엇일까?
- 다양한 시각화 프로그램을 이용하여 텍스트 자료를 어떻게 표현할 수 있을까?

교과서 연계 방안

- 교과서를 통해 텍스트 자료를 벡터로 표현하는 방법을 충분히 이해한 후 활동할 수 있도록 한다.



관련 교과 역량

- | | |
|-----------|--|
| ■ 의사소통 | • 텍스트 자료를 대수적으로 표현하는 방법을 이해할 수 있는가? |
| ■ 태도 및 실천 | • 자료를 수 또는 벡터로 표현하는 것에 어려움을 겪고 있는 모둠원을 배려하고 존중할 수 있는가? |

평가 결과 피드백

- 텍스트 자료를 수나 기호로 표현하고, 다양한 의미의 빈도수를 계산하는 방법의 의미를 정확하게 이해하고 있는지 확인하기 위해 설명하거나 토론하는 활동을 하는 기회를 제공한다.
- 수나 기호로 표현한 자료를 분석하기 위해 시각화하는 과정에서 어떤 수학적 원리를 이용했는지 이해하고 있는지 확인한다.

3. 모듈③

이미지 자료는 어떻게 표현하고 처리할까?

가 모듈 개관

성취기준	12인수02-03 수와 수학 기호를 이용하여 실생활의 이미지 자료를 목적에 알맞게 표현할 수 있다. 12인수02-04 수와 수학 기호로 표현된 이미지 자료를 처리하는 수학 원리를 이해한다.
모듈 설계 의도	성취기준의 의미 인공지능을 이용하면 자료를 정리, 분석하고 패턴을 찾는 수학적 방법을 이해한다. 또한 행렬로 나타낸 이미지 자료를 서로 비교하고 색상을 변형해 봄으로써 이미지 자료를 표현하는 처리하는 수학적 방법을 이해한다. 학습 활동 계획 이 모듈에서는 이미지 자료를 정리, 분석하고 패턴을 찾는 수학적 방법을 이해할 수 있는 탐구 활동을 구성하였다. 기대하는 학습 결과 모듈별 탐구 학습을 통해 인공지능이 이미지 자료를 표현하고 처리하는 방법을 이해할 수 있다. 더 나아가 인공지능에서의 수학의 활용과 중요성을 이해하면서 수학에 대한 가치를 알고 흥미를 가질 수 있다. 모듈 자료에 대한 유의점 인공지능이 이미지를 표현하고 처리하는 수학적 방법을 익히도록 구성되어 있으며, 수학적 개념의 계산보다 이해와 적용에 초점을 두고 있다.
교과 역량	의사소통, 문제 해결, 정보 처리, 태도 및 실천
학습 요소	행렬
교수·학습 방법	탐구 학습
교수·학습 방법 및 유의 사항	• 이미지 자료는 각 픽셀의 위치를 나타내는 가로, 세로 좌표와 색깔을 나타내는 정보로 구성됨을 이해하게 한다. • 이미지의 구도, 색상, 휘도, 밝기, 선명도 등을 간단한 연산을 통해 변경하는 활동을 할 때, 공학적 도구를 이용할 수 있다. • 인공지능과 관련된 수학에 대한 평가를 할 때는 학생들의 흥미, 자신감, 수학의 가치 인식, 학습 태도 등 정의적 영역의 평가와 과정 중심 평가를 할 수 있다.

평가 방법		관찰 평가, 포트폴리오 평가	
평가 방법 및 유의 사항		인공지능에 수학이 어떻게 활용되는지 이해하는 것에 초점을 두어 평가한다.	
차시	주제	교수·학습 활동	평가 활동
1	이미지 자료의 검색	탐구 학습 이미지 검색 프로그램의 이해 - 꽃검색 서비스 알아보기 - 오토드로우 알아보기 - 인공지능의 이미지 인식 기능 이해하기	관찰 평가 인공지능의 이미지 인식 기능을 활용한 앱을 사용할 수 있는가?
2	이미지 자료의 표현	탐구 학습 텍스트 시각화 - 픽셀에 대한 이해하기 - 픽셀의 행렬 표현 이해하기 - 인공지능의 이미지 표현 방법 이해하기	포트폴리오 평가 픽셀의 행렬 표현을 이해하고 표현할 수 있는가?
3	이미지 자료의 변형	탐구 학습 이미지 자료의 변형 - 행렬의 덧셈과 뺄셈 알아보기 - 이미지의 색 변형 알아보기	포트폴리오 평가 행렬의 연산을 이용하여 이미지의 색상을 변형할 수 있는가?

1차시 | 이미지 자료의 검색

학습 목표	다양한 이미지 검색 프로그램을 활용하여 이미지 자료를 검색할 수 있다.
수업의 의도	수업 의도 앱을 활용해 인공지능에서 사용하고 있는 이미지 검색을 경험할 수 있도록 설계하였다. 수업의 주안점 다음(Daum) 앱의 꽃검색 서비스와 오토드로우 프로그램을 이용하여 이미지 검색을 경험할 수 있도록 한다. 수업 활동 시 유의점 인공지능에서 사용하는 이미지 검색이 현재 어느 정도 수준인지 경험하고 그 한계를 찾을 수 있도록 구성되었으며, 앱을 이용하여 다양한 한계를 고려한 의사 결정 시 용어에 어려움을 느낄 수 있다. 의미를 충분히 설명하고 필요시 예시를 들어 학생들이 쉽게 표를 채우고 계산보단 의사 결정에 초점을 맞출 수 있도록 한다. 평가의 주안점 복잡한 계산은 컴퓨터를 활용할 수 있게 하고, 논리적 과정을 중심으로 평가한다.
준비물	활동지, 스마트폰(태블릿 PC로 대체 가능)

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	• 4~5명을 1개 모둠으로 구성	
수행	모둠 활동	• 꽃검색 서비스를 이용한 꽃 이름 검색 활동하기 • 오토드로우를 이용한 단어 맞추기 게임 진행하기	관찰 평가
		• 인공지능의 이미지 인식 기능에 대한 조사하기	
정리		• 인공지능의 이미지 인식 기능에 대해 정리	



이미지 자료의 검색

년

월

일



꽃검색 서비스를 알아보자.

컴퓨터와 수학의 세계에서 사물의 모양을 인식하는 것을 패턴 인식이라고 한다. 다음 그림과 같이 3장의 꽃 사진을 보면 서로 크기도 생김새도 다르지만 ‘장미’를 찍은 사진이다.



사람은 이 사진들을 ‘장미’라고 인식하지만 컴퓨터는 어떻게 인식할 수 있을까? 최근에는 꽃 사진을 찍어 검색해보면 꽃 이름을 바로 알려주는 앱이 있다. 기존의 영상 인식 기술을 활용한 꽃검색 서비스는 정확도가 많이 떨어졌지만, 딥러닝 기술의 발전으로 꽃검색 서비스의 성능이 대폭 향상되었다.

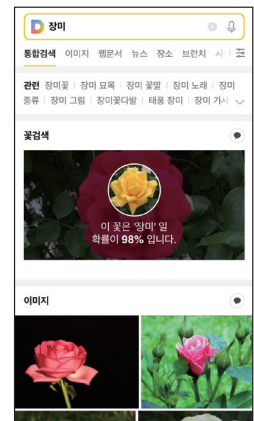
1 다음 순서에 따라 학교 정원에 있는 꽃 또는 나무의 이름을 알아보자.



다음(Daum) 앱에서 검색창의 꽃 모양 아이콘을 누르면 나타나는 꽃 검색을 실행한다.



꽃의 정면을 가운데 안 내선에 맞추어 촬영한다.



꽃 모양 인식 후 나오는 검색 결과를 확인한다.

다음 표 중 (가)에 찍은 사진을 첨부하고, (나)에 꽃의 이름을 적어 보자.

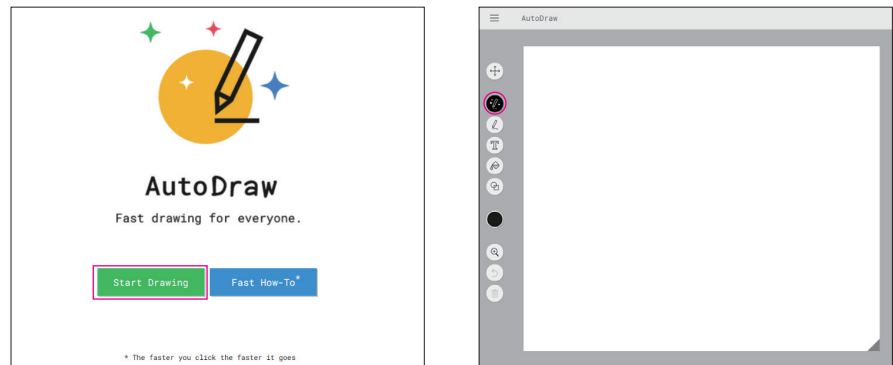
(가)	(나)	(가)	(나)

• 활동팁 주변에 꽃이 없는 경우에는 꽃 사진을 이용해서 확인해 볼 수 있다.

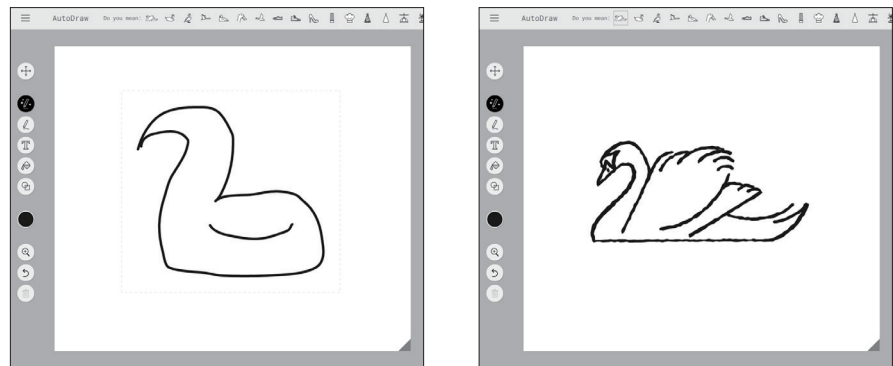
(가)	(나)	(가)	(나)
	꽃잔디		수국
	튤립		코스모스

오토드로우(Auto Draw)를 알아보자.

오토드로우(Auto Draw)는 대략적으로 그린 것을 인공지능이 분석하여 여러 작가가 그린 그림을 추천해주는 프로그램이다. 사용 방법은 다음과 같다.

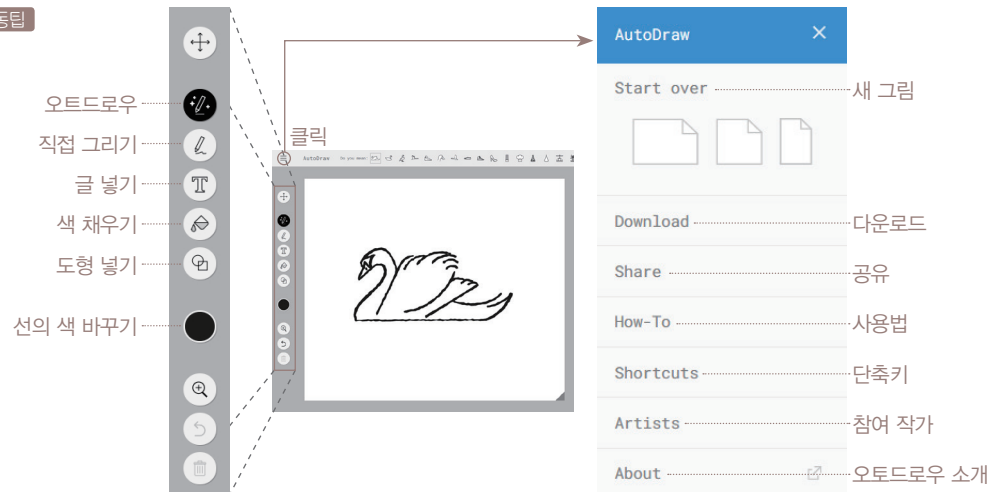


- ① 사이트(<https://www.autodraw.com>)에 접속 후 **Start Drawing**을 누르고 시작한다.
- ② 가장 위의 오토드로우를 누르고 그림을 그린다.



- ③ 그림을 그리면 상단에 추천하는 그림이 나오며, 원하는 그림을 선택할 수 있다.

• 활동팁



1 오토드로우를 이용하여 친구와 단어 맞추기 게임을 해 보자.

- ① 10초 동안 단어에 해당하는 그림을 그린다.
- ② 다 그린 그림을 보고 단어를 유추한다.
- ③ 그린 그림이 상단에 추천하는 그림 중 하나라면 추천한 그림을 눌러 정답을 제시하고, 상단에 추천하는 그림이 아니라면 어떤 그림을 그린 것인지 밝힌다.

• **활동팁** 오토드로우는 발상에 시간이 많이 걸리거나 원하는 그림을 그릴 수 없어 미술을 어려워하고 곤란해 하는 학생에게 도움이 된다. 오토드로우의 도움을 받아 자신의 발상을 구조화하여 이를 직접 그리는 데에 밑그림으로 활용하면 작품을 완성하는 데에 여러 차례 처음부터 다시 그리기를 반복하는 학생들에게 도움이 될 수 있다.

2 인공지능의 이미지 인식 기능이 활용되는 사례를 찾아서 발표해 보자.

인공지능의 이미지 인식은 기계가 마치 사람처럼 사진이나 동영상으로부터 사물을 인식하거나 장면을 이해하는 것을 뜻하며, 이미지 인식은 다음과 같이 3가지가 있다(이주열, 2020).³⁾



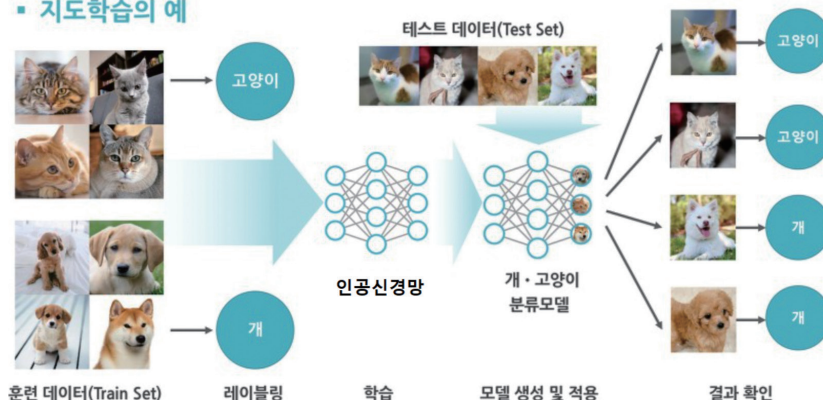
- 분류(Classification): 이미지 내 특정 사물을 분류하는 것
- 검출(Detection): 여러 사물을 동시에 검출하는 것
- 분할(Segmentation): 사물을 픽셀 단위로 식별하는 것

이러한 이미지 인식 기능 중에는 비디오나 이미지에서 얼굴 특징을 매핑하여 인간의 얼굴을 탐지하는 것도 있고, 손동작 인식을 적용하여 가상현실 게임, 수화 등에 사용하기도 하고, 자율주행 자동차에서 운전자의 얼굴을 인식하고 모니터링하여 위험한 주행을 교정하는 적절한 경고 및 개입을 활성화하기도 한다. 최근 미국에서는 택배 물건을 정리해주는 로봇을 만들었다고 한다.

3) 이주열(2020), 인공지능 이미지 인식 기술 동향, TTA저널, 2020 1/2호, p.44-51

인공지능은 사진의 대상을 어떻게 인식할 수 있을까? 첫 번째 방법은 수많은 개와 고양이 사진의 주와 개와 고양이를 분류하게 한다. 이때, 어떤 것이 고양이 사진인지, 개 사진인지 미리 답지를 달아서 알려준다. 훈련 데이터에 답지를 달아주면 인공지능은 개와 고양이의 특징을 추출하여 개와 고양이를 분류하게 한다. 정답이 있는 수많은 데이터를 학습시키는 기술을 지도학습(Supervised Learning)이라고 한다. 답지가 없는 사진을 보고 얼마나 정확하게 개와 고양이를 구분할 수 있는가에 따라 알고리즘의 성능이 결정된다.

■ 지도학습의 예



한편 정답이 없는 수많은 데이터들을 학습시키고 각 데이터들의 특징들을 기계가 스스로 파악하여 서로 비슷한 데이터들을 분류하는 기술을 비지도학습(Unsupervised Learning)이라고 한다. 비지도학습은 답을 맞히는 목적으로 학습하지는 않지만 어떤 데이터들이 서로 비슷한지 그룹지어 주거나, 어떤 성질이 데이터를 잘 정의하는지를 판단하는 등 답을 내지 않는 문제에 대해 나름대로 유용한 정보를 제공하는 역할을 한다. 즉, 비지도학습에서는 구분만 할 뿐 어느 사진이 개인지 고양이인지는 알려주지 않는다. 따라서 그 그룹을 어떻게 정의할지 최종 판단하는 것은 사람의 몫이 된다.

■ 비지도학습의 예



지도의 주안점

- 인공지능의 자료에는 텍스트, 소리, 이미지 등이 있는데 <인공지능 수학>에서는 소리 자료는 다루지 않고, 텍스트와 이미지만을 다루고 있음에 유의한다.
- 인공지능의 이미지 인식 기능을 경험해 보고 현재 어느 수준까지 개발되었는지 확인할 수 있도록 한다.

핵심 발문

- 인공지능의 이미지 인식 기능은 어떤 것이 있을까?
- 인공지능의 이미지 인식 기술을 활용할 때의 유의점은 무엇일까?

교과서 연계 방안

- 이 부분은 교과서에서 다루고 있지 않을 가능성이 있으므로 이미지 자료를 수식으로 표현하는 활동을 하기 전에 활용할 수 있다.



관련 교과 역량

■ 태도 및 실천

- 인공지능의 이미지 검색 기능이 유용하게 활용됨을 이해할 수 있는가?

평가 결과 피드백

- 인공지능의 이미지 검색 기능을 필요에 따라 활용하도록 한다.

2차시 | 이미지 자료의 표현

학습 목표	행렬을 이용하여 이미지 자료를 표현할 수 있다.
수업의 의도	수업 의도 행렬을 이용하여 이미지 자료를 표현하고 픽셀의 행렬 표현을 이해할 수 있도록 설계하였다. 수업의 주안점 각 픽셀을 해당 영역의 색이 표현되는 디지털 정보로 나타낼 때 어떻게 표현할 수 있는지 이해한다. 수업 활동 시 유의점 인공지능에서 사용하는 각 픽셀이 3개의 채널값을 가진다는 것을 이해하고 어떻게 표현되는지 알아보는 활동을 통해 이미지 자료를 행렬로 표현하는 과정을 이해하고 있는지 확인하기 위해 설명하거나 토론하는 활동을 하는 기회를 제공한다. 평가의 주안점 복잡한 계산은 컴퓨터를 활용할 수 있게 하고, 논리적 과정을 중심으로 평가한다.
준비물	활동지, 스마트폰(태블릿 PC, 노트북 등으로 대체 가능)

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	4~5명을 1개 모둠으로 구성	
수행	모둠 활동	- 이미지의 행렬 표현하기 - 각 픽셀을 해당 영역의 색이 표현되는 디지털 정보로 나타낼 때 어떻게 표현할 수 있는지 알아보는 활동하기 - 픽셀의 행렬 표현하기 - 각 픽셀이 3개의 채널값을 가진다는 것을 이해하고 어떻게 표현되는지 알아보는 활동하기	포트폴리오 평가
정리		인공지능의 이미지 표현 방법에 대해 정리	



이미지 자료의 표현

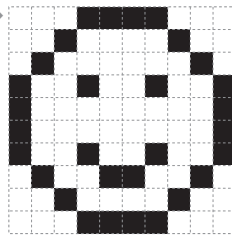
년 월 일



이미지의 행렬 표현을 알아보자.

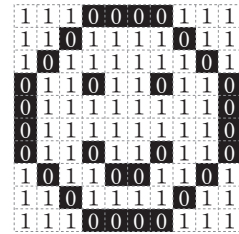
문자, 이미지, 소리와 같이 데이터는 다양하지만, 컴퓨터에 저장할 수 있는 데이터는 결국 숫자들이다. 이 숫자들을 하나의 묶음으로 만들어 의미를 나타낼 수 있는 조합으로 구성할 수 있다.

확소 →



(a)

색상	비트
■	0
□	1



(b)

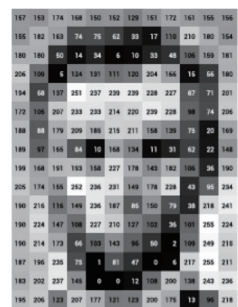
위 그림에서는 흰색과 검은색 2가지 정보로만 구성되므로 각 픽셀의 값은 2가지 경우의 수로 표현할 수 있다. 흰색을 1, 검은색을 0이라고 하면 (b)와 같이 표현할 수 있다. 즉, 각 픽셀이 해당 영역의 색을 나타내는 디지털 정보로 표현되는 것이다. 이때, 해당 픽셀에 검은색이 온전히 채워지지 않고 일부라도 포함되어 있으면 검은색(0)이라고 판단한다.

1 [그림 1]은 해상도가 12×16 인 이미
지이고, [그림 2]는 [그림 1]의 각 픽셀의
자리에 숫자를 하나씩 써넣은 것이다. [그
림 2]에 배열된 수들을 괄호로 묶어 나타낸
행렬을 A라 할 때, 행렬 A는 $m \times n$ 행렬
이다. m, n 의 값을 구해 보자.⁴⁾

$m=12, n=16$



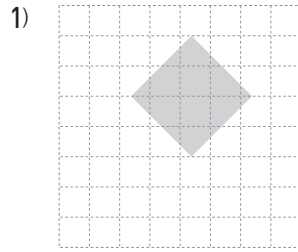
[그림 1]



[그림 2]

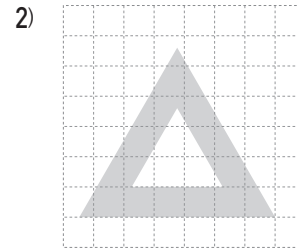
4) <http://assets.runemadsen.com/classes/programming-design-systems/pixels/index.html>

2 다음 그림에서 해당 영역에 색이 포함되어 있으면 1, 그렇지 않으면 0으로 표현하여 행렬로 나타내어 보자.



1)

0	0	0	0	0	0	0	0
0	0	0	1	1	1	0	0
0	0	1	1	1	1	1	0
0	0	1	1	1	1	1	0
0	0	0	1	1	1	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

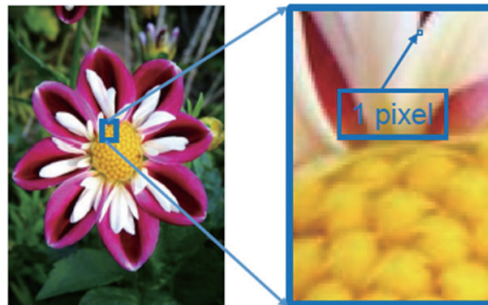


2)

0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0
0	0	1	1	1	0	0	0
0	0	1	1	1	1	0	0
0	1	1	1	1	1	0	0
0	1	1	0	1	1	1	0
1	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0

픽셀과 이미지를 알아보자.

컴퓨터 화면을 확대해보면 작은 점으로 이루어진 것을 볼 수 있다. 이 점은 화면을 구성하는 최소 단위로 픽셀이라고 하며, 작은 정사각형 모양의 각 픽셀에 색상 정보를 이용하여 이미지를 표현할 수 있다.



직사각형 모양의 이미지에 대하여 가로가 a 개의 픽셀, 세로가 b 개의 픽셀로 이루어져 있을 때, 이 이미지의 해상도는 $a \times b$ 라고 한다. 숫자가 클수록 해상도는 높다. 이미지는 수많은 픽셀로 이루어져 있으며, 각각의 픽셀은 빨강(R), 초록(G), 파랑(B)

의 3가지 색상 채널을 가지고 있다. 채널의 양은 0부터 255까지의 정수로 256단계로 표현되며 배합량이 각기 다른 세 개의 원색 채널이 빛의 원리에 따라 혼합되어 특정한 색을 띤다. 채널의 양을 채널값이라고 한다.

예를 들어, 빨강 채널의 채널값이 255, 초록 채널의 채널값이 0, 파랑 채널의 채널값이 0인 경우, $R=255$, $G=0$, $B=0$ 과 같이 나타내고, 이때의 색상은 오른쪽의 그림과 같다.

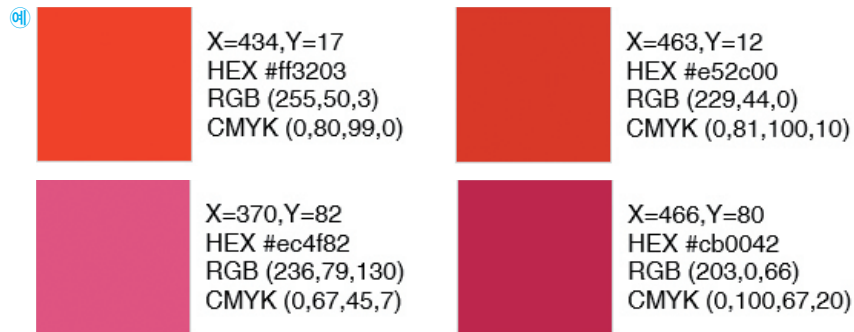
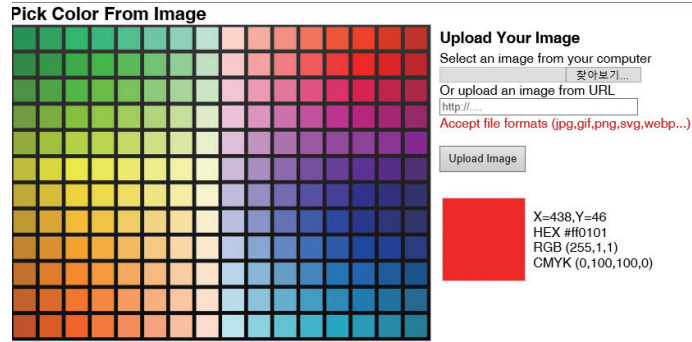
이를 (255, 0, 0)과 같은 순서쌍으로 표현할 수 있으며, (0, 0, 0)은 완전한 검정색이고, (128, 128, 128)은 중간 회색, (255, 255, 255)는 완전한 흰색이다.



* 채널값

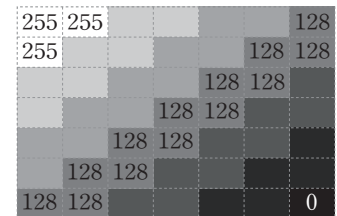
빛의 3원색의 3가지 채널로 색상 정보를 0부터 255까지 정수로 단계적으로 표현할 때, 색상을 나타내는 값

1 사이트(https://www.ginifab.com/feeds/pms/color_picker_image.php)에 접속해서 색을 누르면 채널값을 찾을 수 있다. 좋아하는 색상의 채널값을 찾아보고 같은 빨간 색이라 하더라도 채널값이 어떻게 달라지는지 확인해 보자.



빨간색 계열의 색상은 R값은 크지만 G값 또는 B값의 경우는 100보다 작거나 0인 경우도 있음을 알 수 있다.

2 오른쪽 그림은 흰색부터 검은색까지 단계적으로 명암을 넣어 색칠한 것이다. 흰색을 255, 검은색을 0이라고 할 때 가운데 회색의 값은 128이 된다. 각 단계에 해당하는 색의 값을 구하고, 그 이유를 설명하여 보자.



흰색이 255, 검은색 0이므로 흰색에 가장 가까운 회색을 1단계 회색이라고 할 때, 회색은 총 5단계로 나누어진다. 현재 3단계 회색의 값이 128이므로 1단계 회색의 값은 흰색과 검은색을 5:1로 하는 값이므로

$$\frac{255 \times 5 + 0}{6} = 212.50 \text{이고 반올림하면 } 2130 \text{이므로 } 2130 \text{이 된다. 따라서 2단계 회색의 값은}$$

$$\frac{255 \times 4 + 0}{6} = 170, 4단계 회색의 값은 \frac{255 \times 2 + 0}{6} = 85, 5단계 회색의 값은 \frac{255 \times 1 + 0}{6} = 42.50 \text{이므로}$$

반올림하면 43이 된다.

지도의 주안점

- 이미지를 행렬로 나타낼 수 있음을 이해하게 한다.
- 이미지 자료의 색이 어떻게 결정되는지 이해하게 한다.

핵심 발문

- 이미지를 표현하는 과정에 활용되는 수학적 개념은 무엇일까?
- 이미지의 색은 어떻게 표현될까?
- 행렬의 크기에 따라 이미지는 어떻게 나타날까?

교과서 연계 방안

- 행렬의 의미와 행렬의 표현은 교과서의 내용과 연계하여 사용할 수 있다.



관련 교과 역량

■ 태도 및 실천

- 인공지능의 이미지 자료를 표현하는 방법에 수학이 유용하게 활용됨을 이해할 수 있는가?

평가 결과 피드백

- 이미지 자료를 행렬로 표현하는 과정을 이해하고 있는지 확인하기 위해 설명하거나 토론하는 활동을 하는 기회를 제공한다.
- 이미지 자료의 색이 결정되는 과정에서 어떤 수학적인 원리들이 이용되는지 이해하고 있는지 확인한다.

3차시 | 이미지 자료의 변형

학습 목표	수식을 이용하여 이미지 자료를 변형할 수 있다.
수업의 의도	수업 의도 행렬의 연산을 통해 이미지 자료를 변형할 수 있음을 활동을 통해 이해할 수 있다. 수업의 주안점 행렬의 연산을 이용하여 이미지의 색이 덧칠해질 수 있고, 색이 없어질 수 있음을 이해할 수 있는 활동을 제시하였다. 또한, 기존 이미지를 표현하는 픽셀의 <i>RGB</i>가 각각 행렬로 표현될 때, 색의 명암 및 그레이 스케일로 변형하는 방법을 이해할 수 있는 활동을 제시하였다. 수업 활동 시 유의점 인공지능에서는 수많은 자료를 입력하는 대신 이미지 변형을 통해 새로운 자료를 입력하기도 하는데 이때 이미지 변형은 기존 자료를 행렬로 나타낸 후 행렬의 연산을 통해 변형시킬 수 있다. 변형 과정이 행렬의 연산으로 어떻게 표현할 수 있는지 이해하는 데 초점을 맞출 수 있도록 한다. 평가의 주안점 복잡한 계산은 컴퓨터를 활용할 수 있게 하고, 논리적 과정을 중심으로 평가한다.
준비물	활동지, 스마트폰(태블릿 PC로 대체 가능)

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	4~5명을 1개 모둠으로 구성	
수행	모둠 활동	- 행렬의 덧셈과 뺄셈을 활용한 이미지의 변형하기 - 행렬의 덧셈과 뺄셈을 이용하여 이미지를 더하고 빼는 활동하기 - 이미지의 색상 변형하기 - 세 개의 채널값을 더하고 빼는 활동을 통해 이미지의 색상이 어떻게 변하는지 이해하는 활동	포트폴리오 평가
정리		인공지능의 이미지 변형 방법에 대해 정리	



이미지 자료의 변형

년

월

일



행렬의 덧셈과 뺄셈을 알아보자.

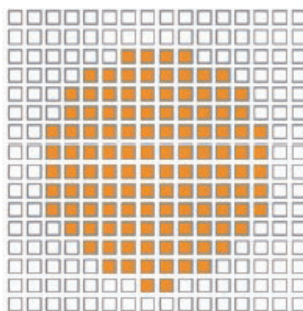
두 행렬 A, B 가 같은 꼴일 때, 두 행렬 A, B 의 대응하는 성분의 합을 성분으로 하는 행렬을 두 행렬 A, B 의 합이라 하고, $A+B$ 와 같이 나타낸다.

한편, 행렬 A 의 각 성분에서 행렬 B 의 대응하는 성분을 뺀 값을 성분으로 하는 행렬을 A 에서 B 를 뺀 행렬이라 하고, $A-B$ 와 같이 나타낸다.

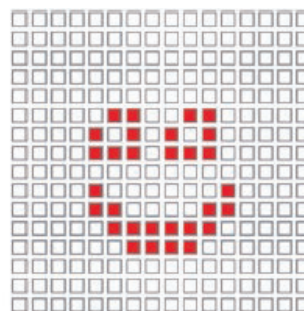
예를 들어, $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ 일 때,

$$A+B = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{pmatrix}, A-B = \begin{pmatrix} a_{11}-b_{11} & a_{12}-b_{12} \\ a_{21}-b_{21} & a_{22}-b_{22} \end{pmatrix} \text{이다.}$$

1 [그림 1]과 [그림 2]의 이미지를 각각 16×16 행렬로 나타내 보고, 두 행렬의 덧셈과 뺄셈을 구해 보자. (단, 주황색은 1, 빨간색은 2, 흰색은 0으로 나타낸다.)



[그림 1]

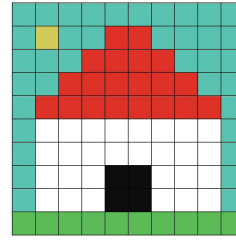


[그림 2]



이미지의 변형을 알아보자.

해상도가 $m \times n$ 인 이미지는 컴퓨터에서 3개의 $m \times n$ 행렬 I_R, I_G, I_B 의 묶음으로 처리된다.



[원본 이미지]

I_R 행렬의 (i, j) 성분은 해당 위치 픽셀의 빨강 채널의 채널값,

I_G 행렬의 (i, j) 성분은 해당 위치 픽셀의 초록 채널의 채널값,

I_B 행렬의 (i, j) 성분은 해당 위치 픽셀의 파랑 채널의 채널값이 되며, 원본 이미지는 다음과 같이 3개의 행렬이 하나의 묶음으로 이루어진다.

67	67	67	67	67	67	67	67	67	67
67	223	67	67	234	234	67	67	67	67
67	67	67	234	234	234	234	67	67	67
67	67	234	234	234	234	234	234	67	67
67	234	234	234	234	234	234	234	234	67
67	255	255	255	255	255	255	255	255	67
67	255	255	255	255	255	255	255	255	67
67	255	255	255	255	255	255	255	255	67
67	255	255	255	255	255	255	255	255	67
60	60	60	60	60	60	60	60	60	60

[빨강 채널]

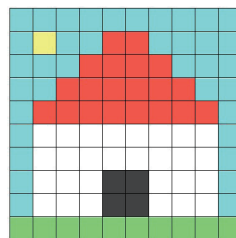
223	223	223	223	223	223	223	223	223	223
223	218	223	223	33	33	223	223	223	223
223	223	223	33	33	33	33	223	223	223
223	223	33	33	33	33	33	33	223	223
223	33	33	33	33	33	33	33	33	223
223	255	255	255	255	255	255	255	255	223
223	255	255	255	255	255	255	255	255	223
223	255	255	255	255	255	255	255	255	223
223	255	255	255	255	255	255	255	255	223
211	211	211	211	211	211	211	211	211	211

[초록 채널]

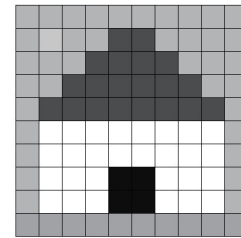
197	197	197	197	197	197	197	197	197	197
197	67	197	197	11	11	197	197	197	197
197	197	197	11	11	11	11	197	197	197
197	197	11	11	11	11	11	11	197	197
197	11	11	11	11	11	11	11	11	197
197	255	255	255	255	255	255	255	255	197
197	255	255	255	255	255	255	255	255	197
197	255	255	255	255	255	255	255	255	197
197	255	255	255	255	255	255	255	255	197
65	65	65	65	65	65	65	65	65	65

[파랑 채널]

또한 원본 이미지에서 세 채널의 모든 채널값에서 50을 더하면 [그림 1]과 같이 원본 이미지보다 밝은 색이 되고 빨강의 각 채널값에 0.2126을 곱하고 초록의 각 채널값에 0.7152를 곱하고 파랑의 각 채널값에 0.0722를 곱한 값을 빨강, 초록, 파랑의 채널값으로 정하면 [그림 2]와 같은 회색 색조의 그림이 된다. 색이 있는 그림을 흑백으로 바꾸려면 빨강, 초록, 파랑의 각 채널값을 동일하게 정하면 된다.



[그림 1]



[그림 2]

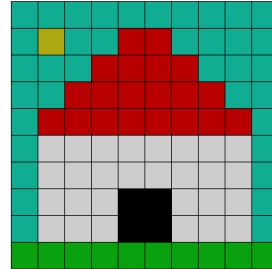


1 해상도가 $m \times n$ 인 이미지는 컴퓨터에서 3개의 $m \times n$ 행렬 I_R, I_G, I_B 로 표시될 때, [그림 1]과 [그림 2]를 만드는 과정을 행렬의 연산으로 표시하여 보자. (단, E 는 모든 성분이 1인 $m \times n$ 행렬이다.)

[그림 1]은 각 채널값에 50을 더한 값이므로 $I_R + 50E, I_G + 50E, I_B + 50E$ 로 나타낼 수 있다.

[그림 2]는 각 채널값을 모두 더하여 3으로 나눈 값이므로 모든 채널값은 $\frac{I_R + I_G + I_B}{3}$ 로 나타낼 수 있다.

반면, [그림 1]의 채널값을 $I_R - 50E, I_G - 50E, I_B - 50E$ 로 바꾸면 [그림 3]과 같이 진하게 표현할 수 있다.



[그림 3]



(AI 웹실험실)의 '이미지 분석'에서 더 많은 실험을 해볼 수 있다.

지도의 주안점

- 이미지를 변형하는 과정에 행렬의 덧셈과 뺄셈, 실수배가 활용됨을 이해할 수 있도록 한다.
- 활동지에는 행렬의 연산에 대한 내용을 제시하지 않았으므로 이 부분은 별도로 학습시킨다.

핵심 발문

- 이미지를 변형하는 과정에 활용하는 수학적 개념은 무엇일까?
- 행렬의 연산에 따라 이미지는 어떻게 변형될까?

교과서 연계 방안

- 행렬의 연산이 이미지의 변형에 영향을 준다는 것을 이해한 후, 다양한 행렬의 연산을 할 수 있도록 교과서의 문제를 활용한다.



관련 교과 역량

- 문제 해결
 - 태도 및 실천
- 다양한 행렬의 연산을 계산할 수 있는가?
 - 인공지능의 이미지 자료를 변형하는 방법에 수학이 유용하게 활용됨을 이해할 수 있는가?

평가 결과 피드백

- 이미지 자료를 행렬로 표현하는 과정을 이해하고 있는지 확인하기 위해 설명하거나 토론하는 활동을 하는 기회를 제공한다.
- 이미지 자료의 색이 결정되는 과정에서 어떤 수학적 원리들이 이용되는지 이해하고 있는지 확인한다.

III

분류와 예측

1. 영역의 개관

자료를 정리, 분석하여 자료의 패턴을 찾을 수 있고 관련된 사건이 일어날 가능성을 수치화할 수 있다. 인공지능을 이용하여 자료를 판별, 정리, 분석하고 패턴을 찾아 새로운 대상에 대한 분류와 예측을 수행하며, 이때 확률과 자료의 시각화, 함수 등을 활용할 수 있다.

가 2015 개정 수학과 교육과정 내용

1. 성취기준 및 학습 요소

분류와 예측 영역에 제시된 성취기준 및 학습 요소는 다음과 같다.

1 자료의 분류

[12인수03-01] 인공지능을 이용하여 텍스트를 분류하는 수학적 방법을 이해한다.

[12인수03-02] 인공지능을 이용하여 이미지를 분류하는 수학적 방법을 이해한다.

2 경향성과 예측

[12인수03-03] 자료를 분석하여 사건이 일어날 확률을 구하고 예측에 이용할 수 있다.

[12인수03-04] 자료의 경향성을 추세선으로 나타내고, 예측에 이용할 수 있다.

관련 학습 요소

- 유사도, 추세선, 조건부확률

각각의 학습 요소를 간단히 설명하면 다음과 같다.

유사도

어떤 대상의 동일함 정도를 수치화한 것으로, 텍스트를 분류하는 수학적 방법인 코사인 유사도 분석은 두 텍스트 사이의 유사 정도를 수치로 나타내고 분석하는 방법을 뜻한다.

추세선

어떤 흐름을 선으로 표시한 것이다.

조건부확률

어떤 사건이 일어났다는 전제 아래에 다른 사건이 일어나는 확률을 뜻한다.

2. 교수·학습 방법 및 유의 사항

분류와 예측 영역의 교수·학습 방법 및 유의 사항은 다음과 같다.

- 텍스트 판별에서는 영화 리뷰 분류, 기사 분류 등을 다룰 수 있고, 글 자료 사이의 유사도를 계산하고 텍스트를 판별하여 분류하는 수학적 과정을 이해하게 한다.

텍스트 자료를 분류하는 분석 방법을 이용하는 대표적인 텍스트 분석 분야인 감성 분석(sentiment analysis)을 주제로 텍스트 자료의 분류 문제를 다룰 수 있다. 텍스트의 긍정/부정/중립 단어의 비율 등을 벡터로 나타내고 이를 이용하여 텍스트 자료를 분류하는 규칙을 학습자가 스스로 제안하고 간단한 텍스트 예시를 분석하여 볼 수 있다. 또한 이러한 기계적 판단의 결과와 인간의 주관적 판단과 비교 평가해 보는 활동을 할 수 있다.

- 이미지 판별에서는 개와 고양이의 사진을 구별하는 문제 등을 다룰 수 있고, 구체적인 이미지 분류 문제의 해결을 위해 인공지능망과 같은 방법이 개발되었음을 인식하게 한다.

개와 고양이의 사진을 구별하거나, 손글씨를 해석하는 문제, 지문 인식, 얼굴 인식 등의 컴퓨터 시각 인식의 문제 등이 이미지를 판별하고 분류하는 인공지능의 대표적인 문제임을 소개하고, 이를 위해 최근 인공지능망을 이용한 딥러닝 등의 기술이 적용됨을 설명한다. 인공지능망은 인간의 신경세포의 정보 전달체계를 수학적으로 모델링한 것임을 설명하고, 인공지능망의 수학적 모델의 구조와 신경망의 유사성을 살펴보게 한다.

- 분류에서는 문자, 음성, 이미지, 안면 등의 인식 시스템이나 문서 유사도를 이용한 표절 검사, 스팸 메일 분류 등의 사례를 다룰 수 있다.

자료의 유형이나 목적에 따라 텍스트를 집합이나 벡터로 나타낸 후 벡터 사이의 거리를 구하거나 코사인 값(코사인 유사도), 자카드 유사도(집합 표현 이용) 등으로 나타내고 비교해 볼 수 있다. 이미지를 분류하는 방법의 예시로 0~9의 숫자의 손글씨 데이터를 이용해 글자 인식의 인공지능을 구현하는 MNIST 데이터의 사례와 함께 손글씨 숫자를 격자로 쪼개 흰색과 검은색을 각각 0과 1로 표현한 행렬로 나타냄으로써, 새로운 손글씨 숫자가 입력되었을 때 이와 제일 가까운 숫자 이미지와의 유사도를 구하여 판별해보는 활동을 해볼 수 있다. 0과 1로 표현된 벡터나 행렬의 유사도를 위해 해밍 거리(hamming distance)나 다른 유사도를 사용할 수 있다.

- 확률은 가능성을 예측하고 싶은 사건을 설정하고 자료를 수집하여 그 사건이 일어날 확률을 구하며, 조건부확률의 용어와 기호를 도입하지 않고 특정 조건으로 세분화된 확률을 상대도수로 추정하고 예측하는 과정에 어떤 수학적 원리가 사용되는지 경험하는 정도로 간단히 다룬다.

자료로부터 확률을 구하는 활동을 통해 인공지능이 수행하는 예측에 확률이 사용됨을 이해하게 한다. 인공지능에서는 분류 결과의 확실성의 정도를 확률을 이용하여 표현함을 알게 한다. 자료로부터 질병이 발생할 확률, 범죄가 일어날 확률 등의 통계적 확률을 구하여 예측하는 활동을 할 수 있다. 또한 조건부확률의 비형식적 도입으로 자료에 나타난 여러 조건에 따른 확률을 구하여 비교해 보거나, 비율 개념을 이용하여 간단한 계산으로 확률이나 기대되는 수치를 추정(기댓값, 모집단에 대한 통계적 추정의 비형식적 도입)하는 등 확률을 통한 예측 활동 및 비형식적인 통계적 추론을 수행한다.

- 자료의 경향성은 관련된 두 자료를 산점도로 나타내고 측정값과 합숫값 사이의 오차를 최소화하는 추세선을 나타내는 두 변량 사이의 관계식 $y=ax+b$ 를 찾아 새로운 x 의 값이 주어졌을 때, y 의 값을 예측하는 과정을 이해하게 한다.

자료를 산점도로 나타내고 상관관계가 있을 때 공학적 도구를 이용하여 자료의 경향성을 보여 주는 일차함수 형태의 추세선을 찾아 새로운 x 의 값에 해당하는 y 의 값을 예측하는 활동을 할 수 있다.

- 추세선은 y 의 측정값과 예측값 사이의 오차를 구하여 이를 최소화한다는 의미 정도로 간단히 다루고, 추세선을 구할 때 공학적 도구를 이용할 수 있다.

회귀분석의 용어는 도입하지 않으며 추세선을 찾는 최소제곱법의 원리는 ‘최적화’에서 다루도록 한다. 공학도구를 사용하여 구한 추세선의 식을 예측에 이용할 수 있음을 중점적으로 다룬다. 또한 자료의 경향성은 직선의 형태가 아닌 경우도 있으므로, 곡선의 경향성을 보이는 자료를 다루어 볼 수도 있다.

- 예측에서는 자동 동작 인식 시스템, 자연어 인식 및 생성 시스템, 자동번역 시스템, 자율주행 자동차, 인공지능 비서, 구매 추천 시스템 등을 다룰 수 있다.

예측에서 활용할 수 있는 인공지능 관련 사례로 자동 동작 인식 시스템, 자연어 인식 및 생성 시스템, 자동번역 시스템, 자율주행 자동차, 인공지능 비서, 구매 추천 시스템 등을 다룰 수 있으며, 이외에도 인공지능과 관련하여 실생활에서 자주 찾을 수 있는 사례를 다룰 수 있다.

3. 평가 방법 및 유의 사항

분류와 예측 영역의 평가 방법 및 유의 사항은 다음과 같다.

- 자료의 예측에 대한 평가는 확률 계산이나 관계식 등 수치 계산에 치중하지 않도록 하고, 자료의 경향성을 추세선으로 나타내는 활동과 확률을 예측에 이용하는 과정에 중점을 둔다.

인공지능의 활용 중 분류와 예측에서 수학이 활용되는 예들을 통해 수학에 대한 가치와 유용성을 느낄 수 있다. 예측에서 활용되는 인공지능의 관련 사례를 통해 자료의 경향성을 나타내는 활동을 공학적 도구를 이용하여 직관적으로 이해할 수 있도록 하는 정의적 평가를 할 수 있다.

2. 모듈④

텍스트 자료는 어떻게 분류할까?

가 모듈 개관

성취기준	12인수03-01 인공지능을 이용하여 텍스트를 분류하는 수학적 방법을 이해한다.
모듈 설계 의도	성취기준의 의미 댓글, 기사 등의 텍스트 자료를 나타내는 벡터를 비교하고 벡터 사이의 거리를 구하여 분류해 봄으로써 인공지능이 텍스트를 분류하는 수학적 방법을 이해한다. 학습 활동 계획 이 모듈은 벡터로 표현된 자료가 인공지능에서 어떻게 활용되는지를 이해하기 위한 활동으로 이루어져 있다. 1차시에서는 코사인 유사도 및 거리 개념을 학습하고 이를 바탕으로 2차시에서는 기사와 같은 텍스트 자료를 카테고리별로 분류해 보거나, 선호도 분류, 감성 분석을 통해 리뷰 분류, 스팸 메일 분류 등의 방법을 이해하고 적용해 보는 학습자 중심으로 탐구할 수 있게 활동이 구성된다. 기대하는 학습 결과 벡터로 표현된 자료를 활용하는 다양한 방법을 배우고 이를 토대로 자료를 분류하고 그 과정을 설명할 수 있다. 모듈 자료에 대한 유의점 학생들이 다양한 텍스트 자료를 활용할 수 있도록 하며, 상황에 따라 다양한 방법으로 문서를 다루어야 함을 이해하게 한 후에 탐구를 진행할 수 있도록 한다.
교과 역량	추론, 의사소통, 정보 처리, 태도 및 실천
학습 요소	벡터, 유사도
교수·학습 방법	탐구 학습
교수·학습 방법 및 유의 사항	• 텍스트 판별에서는 영화 리뷰 분류, 기사 분류 등을 다룰 수 있고, 글 자료 사이의 유사도를 계산하고 텍스트를 판별하여 분류하는 수학적 과정을 이해하게 한다. • 분류에서는 문자, 음성, 이미지, 안면 등의 인식 시스템이나 문서 유사도를 이용한 표절 검사, 스팸 메일 분류 등의 사례를 다룰 수 있다.
평가 방법	포트폴리오 평가
평가 방법 및 유의 사항	

차시	주제	교수·학습 활동	평가 활동
1	유사도의 측정	탐구 학습 유사도의 측정 • 벡터 사이의 거리 구하기 • 코사인 유사도 구하기 • 문서의 유사도 구하기	포트폴리오 평가 • 벡터 사이의 거리를 구할 수 있는가? • 코사인 유사도를 구할 수 있는가? • 자료 사이의 유사도를 구할 수 있는가?
2	텍스트 자료의 분류	탐구 학습 텍스트 자료의 분류 • 문서 사이의 유사도를 구하고, 카테고리 별로 분류하기 • 유사도를 이용하여 선호도 분류하기 • 감성 분석을 통해 리뷰 분류하기 • 인공지능을 이용한 유사도로 분류 원리를 이해하고 스팸 메일 분류하기	포트폴리오 평가 • 코사인 유사도를 활용하여 텍스트 자료를 분류할 수 있는가? • 코사인 유사도로 선호도를 구할 수 있는가? • 감성 분석의 원리를 이해하고 리뷰를 분류할 수 있는가? • 스팸 메일인지 확인하는 범주형 분류 원리를 이해하였는가?

1 차시 | 유사도의 측정

학습 목표	벡터의 내적과 두 점 사이의 거리를 구하는 방법을 알고, 문서 사이의 유사도를 구하는 방법을 설명할 수 있다.
수업의 의도	수업 의도 유클리드 거리, 코사인 유사도를 이용하여 텍스트 자료 사이의 유사도를 측정할 수 있는 방법을 이해한다. 간단한 문서를 비교하는 과정에서 두 가지 방법의 차이점에 대해 이해할 수 있도록 한다. 수업의 주안점 학생들에게 벡터의 내적이나, 두 점 사이의 거리를 구하는 방법을 충분히 이해시킨 후 유사도의 의미를 이해하도록 한다. 또 학생들이 직접 문서들의 유사도를 여러 가지 방법으로 비교해 보는 활동을 통해 유사도의 의미에 대하여 의사소통할 수 있게 한다. 수업 활동 시 유의점 유사도의 계산은 계산기나 컴퓨터를 도입하되 계산 중심의 활동이 되지 않도록 한다. 평가의 주안점 단순히 유사도의 값을 계산하는 게 아니라 각각의 방법이 어떤 차이가 있고 어떤 의미가 있는지 이해하고 있는지에 중점을 두어 평가한다.
준비물	활동지, 계산기(컴퓨터)

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	• 3인 1조로 모둠 구성 • 활동지 배부	
수행	모둠 활동	• 두 점 사이의 거리 이해하기 • 벡터의 내적, 코사인 유사도 이해하기 • 거리 개념과 코사인 유사도를 이용하여 텍스트 자료 사이의 유사도 구하기	포트폴리오 평가
정리		• 유사도의 원리에 대해 정리 • 자기 평가지 작성	자기 평가



유사도의 측정

년

월

일



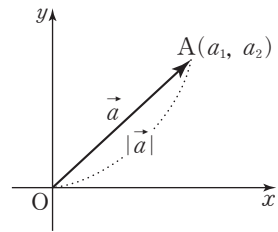
두 점 사이의 거리를 구하는 방법과 코사인 유사도에 대해 알아보자.

1) 두 점 사이의 거리(유클리드 거리, Euclidean distance)

좌표평면에서 두 점 (a_1, a_2) , (b_1, b_2) 사이의 거리는 $\sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}$ 이다. 마찬가지로 n 차원 공간에서 두 점 $(a_1, a_2, \dots, a_n)$, $(b_1, b_2, \dots, b_n)$ 사이의 거리는 좌표평면에서와 같이 $\sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \dots + (b_n - a_n)^2}$ 으로 정의한다.

2) 벡터의 크기와 내적

벡터를 표기하는 방법에 대해 알아보자. 오른쪽 그림과 같이 a_1, a_2 를 성분으로 하는 자료 (a_1, a_2) 는 좌표평면 상의 한 점 A를 나타낸다. 이때, 시작점을 원점 O, 끝점을 A로 하는 화살표로 나타낸 것을 벡터라고 하며 $\overrightarrow{OA} = (a_1, a_2)$ 로 표기하기도 하고, 간단하게 $\vec{a} = (a_1, a_2)$ 로 표기하기도 한다. 벡터를 이루는 각각의 성분은 하나의 숫자로 이루어져 있는데 이를 스칼라라고 한다.

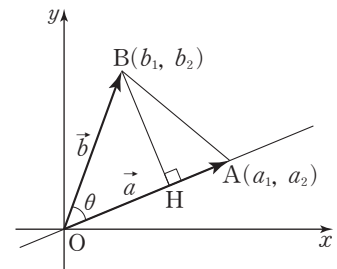


또, 벡터 $\vec{a}$ 의 크기는 $|\vec{a}| = \sqrt{a_1^2 + a_2^2}$ 과 같이 나타낸다. 마찬가지로 $\vec{a} = (a_1, a_2, \dots, a_n)$ 인 경우 벡터의 크기는 $|\vec{a}| = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}$ 이다. 따라서 벡터의 크기는 시작점과 끝점 사이의 거리를 의미한다고 할 수 있다.

이제 벡터의 내적에 대하여 알아보자.

두 벡터 $\vec{a}$, $\vec{b}$ 가 이루는 각의 크기를 θ 라고 하면 $|\vec{a}| |\vec{b}| \cos \theta$ 를 $\vec{a}$ 와 $\vec{b}$ 의 내적이라 하고, 이것을 기호로 $\vec{a} \cdot \vec{b}$ 와 같이 나타낸다.

오른쪽 그림과 같이 영벡터가 아닌 두 벡터 $\overrightarrow{OA} = \vec{a} = (a_1, a_2)$, $\overrightarrow{OB} = \vec{b} = (b_1, b_2)$ 가 이루는 각의 크기를 θ 라고 하자.



삼각형 OAB의 꼭짓점 B에서 직선 OA에 내린 수선의 발을 H라고 하면 두 직각삼각형 HAB와 OHB에서 다음이 성립한다.

$$\begin{aligned}
|\overrightarrow{AB}|^2 &= |\overline{AB}|^2 = |\overline{AH}|^2 + |\overline{BH}|^2 \\
&= (\overline{OA} - \overline{OH})^2 + (\overline{OB}^2 - \overline{OH}^2) \\
&= \overline{OA}^2 + \overline{OB}^2 - 2\overline{OA} \times \overline{OH} \\
&= \overline{OA}^2 + \overline{OB}^2 - 2\overline{OA} \times \overline{OB} \cos \theta \\
&= |\overrightarrow{OA}|^2 + |\overrightarrow{OB}|^2 - 2\overrightarrow{OA} \cdot \overrightarrow{OB}
\end{aligned}$$

이때, 벡터의 크기를 성분으로 나타내면

$$(b_1 - a_1)^2 + (b_2 - a_2)^2 = (a_1^2 + a_2^2) + (b_1^2 + b_2^2) - 2\vec{a} \cdot \vec{b} \text{이고 이 식을 정리하면}$$

$$\vec{a} \cdot \vec{b} = a_1 b_1 + a_2 b_2$$

와 같다. 마찬가지로 n 차원 벡터 $\vec{a} = (a_1, a_2, \dots, a_n)$, $\vec{b} = (b_1, b_2, \dots, b_n)$ 의 내적 $\vec{a} \cdot \vec{b}$ 는 다음과 같이 정의한다.

$$\vec{a} \cdot \vec{b} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$$

두 벡터의 내적의 결과는 실수이므로 일반적인 연산이 가능하다. 따라서 식을 변형하면

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|} = \frac{\vec{a}}{|\vec{a}|} \cdot \frac{\vec{b}}{|\vec{b}|} = \frac{a_1 b_1 + a_2 b_2 + \dots + a_n b_n}{\sqrt{a_1^2 + a_2^2 + \dots + a_n^2} \sqrt{b_1^2 + b_2^2 + \dots + b_n^2}}$$

과 같이 나타낼 수 있다.

3) 코사인 유사도

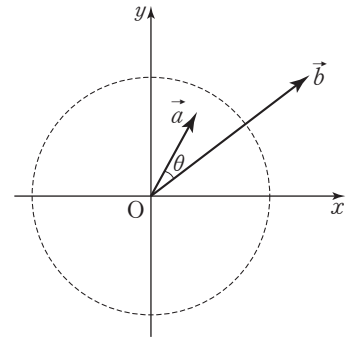
위의 식 $\cos \theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|} = \frac{\vec{a}}{|\vec{a}|} \cdot \frac{\vec{b}}{|\vec{b}|}$ 에서 두 벡터

$\frac{\vec{a}}{|\vec{a}|}$ 와 $\frac{\vec{b}}{|\vec{b}|}$ 는 크기가 1인 벡터이므로 자료 사이

의 거리는 무시하고 자료의 패턴(방향)만을 고려하게 된다. 따라서 어떤 두 자료를 비교할 때, 위의 코사인 값을 통해 두 자료의 유사성을 비교한다.

즉, θ 가 작으면 자료의 유사도가 높고, θ 가 크면 자료의 유사도가 낮다고 판단할 수 있으며 $\cos \theta$ 로

유사도를 측정한다. 이를 코사인 유사도(cosine similarity)라고 한다.



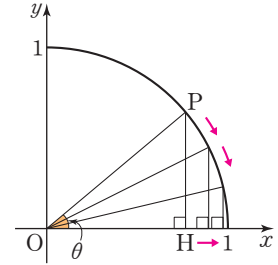
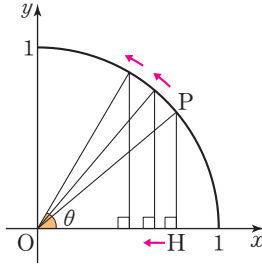
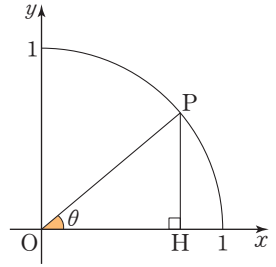
두 자료 $\vec{a} = (a_1, a_2, \dots, a_n)$ 과 $\vec{b} = (b_1, b_2, \dots, b_n)$ 의 코사인 유사도는 다음과 같이 계산할 수 있다.

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|} = \frac{\vec{a}}{|\vec{a}|} \cdot \frac{\vec{b}}{|\vec{b}|} = \frac{a_1 b_1 + a_2 b_2 + \dots + a_n b_n}{\sqrt{a_1^2 + a_2^2 + \dots + a_n^2} \sqrt{b_1^2 + b_2^2 + \dots + b_n^2}}$$

코사인의 값이 커지면 사잇각은 작아지고, 유사도는 높아진다.

참고

그림과 같이 좌표평면 위에 원점 O 를 중심으로 하고 반지름의 길이가 1인 사분원 위의 한 점 P 에서 x 축에 내린 수선의 발을 H 라고 하면, $\cos \theta = \frac{\overline{OH}}{\overline{OP}} = \overline{OH}$ 이므로 θ 의 값이 커지면 $\cos \theta$ 의 값이 작아지고, θ 의 값이 작아지면 $\cos \theta$ 의 값이 커짐을 알 수 있다.



θ 의 값이 커지면 $\cos \theta$ 의 값은 작아진다. θ 의 값이 작아지면 $\cos \theta$ 의 값은 커진다.

1 다음 벡터의 코사인 유사도를 구해 보자.

벡터	$ \vec{a} $	$ \vec{b} $	$\vec{a} \cdot \vec{b}$	코사인 유사도
$\vec{a} = (1, 1), \vec{b} = (4, 3)$	$\sqrt{2}$	5	7	$\frac{7\sqrt{2}}{10}$
$\vec{a} = (-1, 0, 1), \vec{b} = (2, 4, 4)$	$\sqrt{2}$	6	2	$\frac{\sqrt{2}}{6}$

2 두 벡터 $\vec{a} = (1, 1, -1), \vec{b} = (2, 1, 2)$ 에 대하여 각각 유클리드 거리와 코사인 유사도를 구해 보자.

- 유클리드 거리: $\sqrt{(2-1)^2 + (1-1)^2 + (2-(-1))^2} = \sqrt{10}$
- 코사인 유사도: $\frac{2+1-2}{\sqrt{1^2+1^2+(-1)^2}\sqrt{2^2+1^2+2^2}} = \frac{1}{3\sqrt{3}} = \frac{\sqrt{3}}{9}$



다음은 영화 리뷰 문서를 보고 빈도수를 작성한 것이다.

영화	수사	유머	범인	살인	콤비	단서	재벌	악역
탐정: 더 비기닝	1	1	1	0	1	1	0	0
탐정: 리턴즈	3	3	1	0	3	2	0	0
베테랑	2	0	0	1	0	0	1	1

• **활동팁** 문서 행렬을 TF-IDF로 변형한 후에 문서 유사도를 측정할 수도 있다.

1 리뷰 사이의 유클리드 거리(Euclidean distance)를 구해 보자.

	유클리드 거리
탐정: 더 비기닝 & 탐정: 리턴즈	$\sqrt{(1-3)^2 + (1-3)^2 + (1-3)^2 + (1-2)^2} = \sqrt{13} \approx 3.6$
탐정: 더 비기닝 & 베테랑	$\sqrt{(1-2)^2 + (1-0)^2 + (1-0)^2 + (0-1)^2 + (1-0)^2 + (1-0)^2 + (0-1)^2 + (0-1)^2} = \sqrt{8} \approx 2.8$
탐정: 리턴즈 & 베테랑	$\sqrt{(3-2)^2 + (3-0)^2 + (1-0)^2 + (0-1)^2 + (3-0)^2 + (2-0)^2 + (0-1)^2 + (0-1)^2} = 3\sqrt{3} \approx 5.2$

• **활동팁** 각각의 문서를 벡터로 나타내면 '탐정: 더 비기닝'은 (1, 1, 1, 0, 1, 1, 0, 0), '탐정: 리턴즈'는 (3, 3, 1, 0, 3, 2, 0, 0), '베테랑'은 (2, 0, 0, 1, 0, 0, 1, 1)이다.

2 리뷰 사이의 코사인 유사도(Cosine similarity)를 구해 보자.

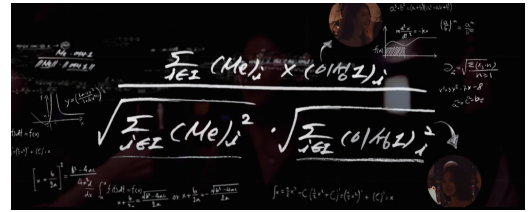
	코사인 유사도
탐정: 더 비기닝 & 탐정: 리턴즈	$\frac{3+3+1+3+2}{\sqrt{1^2+1^2+1^2+1^2+1^2}\sqrt{3^2+3^2+1^2+3^2+2^2}} \approx 0.95$
탐정: 더 비기닝 & 베테랑	$\frac{2}{\sqrt{1^2+1^2+1^2+1^2+1^2}\sqrt{2^2+1^2+1^2+1^2}} \approx 0.34$
탐정: 리턴즈 & 베테랑	$\frac{6}{\sqrt{3^2+3^2+1^2+3^2+2^2}\sqrt{2^2+1^2+1^2+1^2}} \approx 0.4$

3 유클리드 거리와 코사인 유사도의 값을 비교해 보고 그 이유에 대해 생각해 보자.

실제로 '탐정: 더 비기닝'과 '탐정: 리턴즈'가 유사한 영화이지만 유클리드 거리는 가장 크게 나와 유사하지 않은 결과를 얻었다. 긴 문서를 택한 경우 비슷한 유형의 단어가 반복되는 횟수가 많아질 수 있다. 즉, 비슷한 단어가 쓰였지만 많은 단어가 사용되면 유클리드 거리가 커지게 되기 때문이라고 할 수 있다. 반면 코사인 유사도는 벡터의 방향을 측정하는 것으로 벡터의 크기보다는 방향성을 중심으로 비교하기 때문에 '탐정: 더 비기닝'과 '탐정: 리턴즈'가 가장 유사한 것으로 나타났음을 알 수 있다.



다음 영상을 보고 유사도를 활용할 수 있는 방법에 대해 생각해 보자.



(<https://www.youtube.com/watch?v=BpB4pYOY25Q>)

1 AI 웹실험실 을 이용하여 리뷰 유사도를 확인해 보자.

코사인 유사도

index	문장1	문장2	문장3	문장4	문장5	문장6	문장7	문장8	문장9	문장10
문장1	1	0	0.04	0.11	0.18	0.02	0	0.04	0	0
문장2	0	1	0	0	0	0	0	0	0	0.44
문장3	0.04	0	1	0.16	0.23	0.72	0	0.05	0	0
문장4	0.11	0	0.16	1	0.16	0.07	0	0.03	0	0.11
문장5	0.18	0	0.23	0.16	1	0.1	0	0.21	0	0
문장6	0.02	0	0.72	0.07	0.1	1	0.39	0.26	0	0.09
문장7	0	0	0	0	0	0.39	1	0.3	0	0.11
문장8	0.04	0	0.05	0.03	0.21	0.26	0.3	1	0	0
문장9	0	0	0	0	0	0	0	0	0	0
문장10	0	0.44	0	0.11	0	0.09	0.11	0	0	1

[웹실험실의 유사도 계산 예시]

읽을 거리 1

자카드 유사도

자카드 유사도(Jaccard Similarity)는 두 문장 A, B 를 각각 단어 집합으로 만든 뒤 두 집합을 통해 유사도를 측정하는 방식이다. 두 집합의 교집합인 공통된 단어의 개수를 두 집합의 합집합의 단어의 수로 나누면 된다. 즉, 두 문서 A, B 에 대한 자카드 유사도는 $\frac{|A \cap B|}{|A \cup B|}$ 이다.

따라서 결과 값은 공통의 원소의 개수에 따라 0과 1 사이의 값이 나오며 1에 가까울수록 유사도가 높다는 의미로 해석할 수 있다.

예를 들어,

A 문장에 사용된 단어: 컴퓨터, 사용, 인공지능, 유사도, 코사인

B 문장에 사용된 단어: 컴퓨터, 인공지능, 벡터, 문서, 유사도

일때

$A \cup B = \{\text{컴퓨터, 사용, 인공지능, 벡터, 문서, 유사도, 코사인}\}$

$A \cap B = \{\text{컴퓨터, 인공지능, 유사도}\}$

이므로 자카드 유사도는 $\frac{3}{7}$ 이다.

워드투벡터(Word2vec)를 이해하기 전에 다음의 정의를 먼저 살펴보자.

- 1) 원-핫 인코딩(One-Hot Encoding): 단어 집합의 크기를 벡터의 차원으로 하고, 표현하고 싶은 단어의 인덱스에 1의 값을 부여하고, 다른 인덱스에는 0을 부여하는 단어의 벡터 표현 방식이다.

예 '나는', '너를', '사랑해'라는 단어사전이 있는 경우 '나는', '너를', '사랑해'라는 단어는 각각 벡터로 (1, 0, 0), (0, 1, 0), (0, 0, 1)과 같이 표현하는 방법이다.

- 2) 희소 벡터: 원-핫 벡터들은 표현하고자 하는 단어의 인덱스의 값만 1이고, 나머지 인덱스에는 전부 0으로 표현되는 벡터 표현 방법이다. 이때, 벡터의 성분 대부분이 0으로 표현되는 방법을 희소 표현(sparse representation)이라고 하며 원-핫 벡터는 희소 벡터라고 한다.

- 3) 밀집 벡터(dense vector): 희소 표현과 반대로 벡터의 차원을 단어 집합의 크기로 상정하지 않고 일반적으로 줄여서 표현하는 것을 밀집 표현(dense representation)이라고 하며 밀집 표현으로 나타낸 벡터를 밀집 벡터라고 한다.

희소 벡터 (0, 1, 0, 0, 0)을 밀집 벡터 (1.2, 0.42, -3.1)과 같이 차원을 줄여서 표현 가능하다.

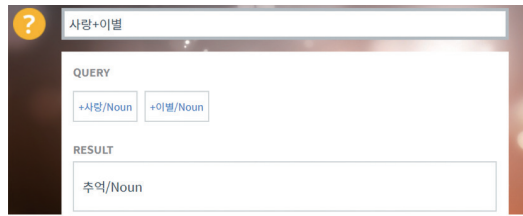
- 4) 워드 임베딩(word embedding): 단어를 밀집 벡터(dense vector)의 형태로 표현하는 방법을 워드 임베딩(word embedding)이라고 한다. 그리고 이 밀집 벡터를 워드 임베딩 과정을 통해 나온 결과라고 하여 임베딩 벡터(embedding vector)라고도 한다.

	원-핫 벡터	임베딩 벡터
차원	고차원(단어 집합의 크기)	저차원
표현 형식	희소 벡터의 형태로 표현	밀집 벡터의 형태로 표현
값의 타입	1과 0	실수
표현 방법	수동(임의로 지정)	훈련 자료로 학습

원-핫 벡터는 단어 간 유사도를 계산할 수 없다. (코사인 유사도 계산 시 내적의 값이 모두 0이다.) 단어 간 유사도를 반영할 수 있도록 단어의 의미를 벡터화할 수 있는 방법이 고안되었으며 그 대표적인 방법이 워드투벡터(Word2Vec)이다.

원-핫 벡터를 학습을 통해 의미 있는 단어 벡터로 만들어 단어 사이에 유사도를 갖도록 한다. 유사도를 갖도록 만든 워드투벡터에서 고양이와 애교를 나타내는 벡터를 더하는 경우 강아지를 나타내는 벡터를 얻기도 하며, 한국, 서울, 도쿄, 일본을 나타내는 벡터에 대하여 '한국-서울+도쿄=일본'과 같은 연산 결과를 얻을 수 있다.

아래 사이트는 한국어 단어에 대한 벡터 연산이 가능하도록 워드투벡터 기반으로 만들었다. 다음 사이트 <https://word2vec.kr/search/>에서 단어 벡터의 연산을 해 볼 수 있다.



또한 워드투벡터로 변환된 단어 사이의 거리를 비교하여 두 문서 사이의 유사도를 구할 수 있다.

지도의 주안점

- 교사는 자료의 유사도를 비교하는 활동을 통해 유사도의 의미를 이해하도록 도움을 준다. 또한 유사도를 계산하는 경우, 컴퓨터 프로그램 등을 적극적으로 활용하도록 한다.

핵심 발문

- 텍스트 자료의 유사도를 구하는 식은 무엇일까?
- 유클리드 거리와 코사인 유사도의 차이점은 무엇일까?

교과서 연계 방안

- 교과서를 통해 벡터, 내적, 유사도의 계산 방법 등을 충분히 이해한 후 활동할 수 있도록 한다.



관련 교과 역량

- 의사소통
- 태도 및 실천

- 텍스트 자료의 유사도를 측정하는 방법을 이해할 수 있는가?
- 자료 사이의 유사도를 구하고 비교하는 활동에 어려움을 겪고 있는 모둠원을 배려하고 존중할 수 있는가?

평가 결과 피드백

- 벡터로 표현된 자료의 유사도를 계산하는 방법의 의미를 정확하게 이해하고 있는지 확인하기 위해 설명하거나 토론하는 활동을 하는 기회를 제공한다.
- 유사도를 이용하여 문서 사이의 유사도를 측정하는 과정에서 어떤 수학적 원리를 이용했는지 이해하고 있는지 확인한다.

2차시 | 텍스트 자료의 분류

학습 목표		벡터의 내적과 두 점 사이의 거리를 구하는 방법을 알고, 문서 사이의 유사도를 구하는 방법을 설명할 수 있다.	
수업의 의도		수업 의도 코사인 유사도를 이용하여 텍스트 자료를 카테고리별로 분류해 보거나 스팸 메일을 분류해 보는 활동을 통해 분류하는 다양한 방법에 대해 이해할 수 있도록 한다.	
		수업의 주안점 카테고리 분류, 손님 사이의 유사도 분류, 스팸 메일 분류 등 다양한 활동을 통해 여러 가지 분류 방법을 생각해 보는 활동을 해 보며 분류의 다양한 원리에 대해 의사소통할 수 있게 한다.	
		수업 활동 시 유의점 유사도의 계산은 계산기나 컴퓨터를 도입하되 활동이 계산 중심이 되지 않게 한다.	
		평가의 주안점 다양한 활동에서 상황에 맞게 적절하게 분류하는 방법에 대해 충분하게 이해하고 있는 지에 중점을 두어 평가한다.	
준비물		활동지, 계산기(컴퓨터)	

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	2~3인 1조로 모둠 구성 - 활동지 배부	
수행	모둠 활동	- 문서 카테고리 분류하기 - 손님 사이의 유사도 구하기 - 감성 분석하기 - 스팸 메일 분류하기	포트폴리오 평가
정리		- 분류 알고리즘에 대해 정리 - 자기 평가지 작성	



텍스트 자료의 분류

년 월 일



글의 주제를 파악하여 적당한 카테고리로 분류해 보자.

매일 다양한 분야에서 논문이 쏟아지고 있다. 이때, 작성된 논문은 해당 분야의 전문가에게 전달되어야 논문 심사가 제대로 이루어질 수 있다. 하지만 굉장히 많은 양의 논문을 다 읽고 분류하는 것은 굉장히 힘든 일이며 따라서 어떤 논문이 어떤 분야에 해당하는 것인지 파악하는 데 걸리는 시간도 상당히 소모될 수 있다. 마찬가지로 카페나 다양한 커뮤니티에 올라오는 다양한 글을 주제별로 분류하는 일은 지금까지 사람이 일일이 해왔지만 점차적으로 인공지능이 그 역할을 대신하고 있다.

그럼 어떻게 문서를 특정 주제에 맞게 분류할 수 있을까? 보통 특정한 카테고리에 속하는 글들은 특유의 문체가 있기도 하고 특정한 단어를 자주 사용하게 마련이다. 예를 들어, 과학에 관한 글이라면 ‘법칙’이나 ‘원리’와 같은 단어가 많이 사용될 것이며, 인문학에 관한 글들은 ‘사상’이나 ‘철학’과 같은 단어가 사용될 것임을 추측할 수 있다. 만약 특정 카테고리의 특징어를 알고 있고, 그러한 특징어가 포함된 글이 주어진다면 해당 글이 어떤 내용인지 조금 더 쉽게 분류할 수 있을 것이다.

1 다음은 ‘인문’, ‘과학’의 카테고리로 분류하여 각각의 문서에 사용된 주요 단어의 빈도수를 나타낸 것이다.

	법칙	사상	가설	입자	지능	심리
과학 분야 문서	6	1	3	1	1	0
인문 분야 문서	2	6	2	0	1	2

위 단어의 순서대로 문서 A와 문서 B를 나타내는 벡터가 (1, 3, 1, 0, 1, 1), (3, 1, 2, 0, 1, 0) 일 때, 문서 A와 문서 B를 과학 분야 또는 인문 분야의 문서로 분류해 보자.

과학 분야, 인문 분야 문서를 나타내는 벡터는 각각 (6, 1, 3, 1, 1, 0), (2, 6, 2, 0, 1, 2)이므로 각각 문서의 코사인 유사도를 구해 보자.

① 과학 분야 문서 & 문서 A: $\frac{13}{\sqrt{48}\sqrt{13}} \approx 0.52$

② 과학 분야 문서 & 문서 B: $\frac{26}{\sqrt{48}\sqrt{15}} \approx 0.97$

③ 인문 분야 문서 & 문서 A: $\frac{25}{\sqrt{49}\sqrt{13}} \approx 0.99$

④ 인문 분야 문서 & 문서 B: $\frac{17}{\sqrt{49}\sqrt{15}} \approx 0.63$

따라서 문서 A는 인문 분야, 문서 B는 과학 분야의 문서로 분류할 수 있다.

2 실제로는 문서의 양도 많고 카테고리 수도 많다. 많은 양의 문서를 카테고리에 분류해야 한다고 가정하자. 어떠한 방법으로 분류하면 좋을지 자유롭게 이야기해 보자.

모든 문서끼리 코사인 유사도를 계산하여 분류할 수 있겠지만 시간이 많이 걸릴 것이다. 또한 어떤 카테고리 분류에 할지도 알 수 없는 상황이다.

문서의 벡터를 이용하여 두 개씩 짝 지은 문서의 코사인 유사도를 구한다. 유사성이 기준 값(예 0.6) 이상이면 소분류로 합친다. 그러면 문서의 개수가 일정 범위 줄어들 것이다.

이때 소분류의 모든 문서를 하나로 보고 벡터를 구하고, 마찬가지로 방법으로 소분류의 코사인 유사도를 이용하여 더 큰 소분류로 합친다.

이외에도 다양한 답이 있을 수 있다.



다음 제시문을 읽고 물음에 답해 보자.

넷플릭스나 교보문고와 같은 사이트에서는 특별한 영상이나 책을 추천하는 서비스를 제공하고 있다. 사용자의 영상 시청 패턴이나 구매 패턴을 분석하여 추천하는 서비스를 통해 구매자에게 다양한 정보를 제공한다.

○○○님을 위한 Picks 추천 샘플 [로그인 하고 나만의 AI 맞춤추천 받아보세요] ?



[출처: www.kyobobook.co.kr]

마찬가지로 인터넷상에서 앱을 통한 주문이 늘어나고 손님에 대한 주문 정보가 쌓이면 이를 바탕으로 손님들의 선호도를 파악하여 손님에게 적당한 음식을 추천해줄 수 있으며 더욱 효율적으로 운영이 가능할 것이다.

[1-3] 아래는 손님별로 음식 선호도(1점~5점)를 나타낸 표이다. (주문하지 않은 경우는 0점)

배달 음식	손님 1	손님 2	손님 3	손님 4
돈가스	5	1	0	1
치킨	4	0	0	4
짜장면	4	0	0	0
피자	0	5	4	4
햄버거	0	0	5	0

1 손님 사이의 코사인 유사도를 구하여 표를 작성해 보자. 음식 선호도에서 가장 유사한 성향을 가지고 있는 손님을 말해 보자.

	손님 1	손님 2	손님 3	손님 4
손님 1	1	0.13	0	0.48
손님 2	0.13	1	0.61	0.72
손님 3	0	0.61	1	0.43
손님 4	0.48	0.72	0.43	1

본인을 제외하면 손님 2와 손님 4의 유사도가 0.72로 가장 높다.

2 손님 2에게 음식을 추천한다면 어떤 음식을 추천할 것인가?

손님 2와 유사한 손님은 손님 4이고 손님 4는 치킨을 먹었으므로 손님 2에게 치킨을 추천한다.

• 활동팁 공통으로 주문한 음식에 대해서만 손님 간의 유사도를 계산할 수도 있다.

3 피자와 유사도가 높은 음식은 어떤 음식이라고 말할 수 있는지 알아보고, 피자를 먹은 손님 2에게 어떤 음식을 추천하면 좋을지 이야기해 보자.

음식에 대하여 각각의 손님을 성분으로 하는 벡터를 생각하자.

즉, 돈가스는 (5, 1, 0, 1), 치킨은 (4, 0, 0, 4), 짜장면은 (4, 0, 0, 0), 피자는 (0, 5, 4, 4), 햄버거는 (0, 0, 5, 0)이라고 할 수 있다. 피자과 다른 음식간의 유사도를 구해 보자.

	돈가스	치킨	짜장면	햄버거
피자	0.23	0.37	0	0.53

따라서 손님 2에게 햄버거를 추천한다.



감성 분석에 대해 알아보자.

감성 분석(Sentiment Analysis)은 특정 문서의 긍정, 부정에 대한 감정을 추측하고 분류하는 방법이다. 여러 가지 방법이 있으나 대표적인 방법 중의 하나는 단어를 긍정, 부정, 중립 등으로 분류한 감성 사전을 구축한 후 주어진 문서에 출현한 단어를 비교하여 문서를 분류하는 것이다. 즉, 영화 리뷰나, 기사의 댓글, 어떤 상품의 후기와 같은 텍스트 자료에서 긍정적인 단어의 출현 빈도가 높다면 긍정적으로 느끼고 있다고 판단하여 만족도 등을 분석할 수 있다.

다음 표는 긍정, 부정 어휘에 대한 감성 사전의 예시이다.

긍정					부정				
같은	보다	대	상	밝은	의	죽은	피해	소수	좁은
더	전체	개인	혁명	전용	한	문제	부인	시험	나쁜
큰	최고	자유	미	크다	다른	사용하여	대상	어두운	노예
주	작품	기능	충분히	상당히	제한	부족한	적	전차	어렵다
같이	현대	초	아름다운	방법	특히	신	잘못	결정	다리를
매우	프로	빠른	구조	빨리	뒤	본래	붙어	죽을	경쟁
잘	사용할	기술	젊은	간단한	없는	제작	아래의	비판	걸려

[출처: <https://www.kaggle.com/rtatman/sentiment-lexicons-for-81-languages>]

1 한국어 감성 사전을 제공하고 있는 사이트(<http://dilab.kunsan.ac.kr/knu/knu.html>)에서 단어별 감성 점수를 확인하여 긍정, 중립, 부정 리뷰로 분류해 보자. 또, 분류한 결과와 실제로 느끼는 감성을 비교해 보자.

리뷰 1: 고급스런 디자인에 가격도 저렴해서 만족한다. ^^

리뷰 2: 따뜻하고 사용하기에 편하지만 가격이 비싸서 부담된다. --;

리뷰 3: 저렴하긴 하지만 디자인이 별로다.

리뷰 1은 고급스런: +2, 저렴해서: +2, 만족한다: +2, ^^: +10이므로 합계는 +7.

리뷰 2는 따뜻하고: +2, 편하지만: +2, 비싸서: -2, 부담된다: -2, --: -10이므로 합계는 -1.

리뷰 3은 저렴하긴: +1, 별로다: -10이므로 합계는 0.

따라서 리뷰 1은 긍정, 리뷰 2는 부정, 리뷰 3은 중립이라고 할 수 있다.

또, 실제 문장에서 느껴지는 감성도 비슷하다고 할 수 있다.

2 다음 단어에 대하여 감성 점수를 부여해 보고, 친구 및 사이트에서 제공하는 점수와 그 결과를 비교해 보자.

단어	부족하다	비슷하다	행복하다	아리송하다	미흡하다
내 점수	-2	0	+2	0	-1
친구의 점수	-1	+1	+2	-1	-2
감성 사전 사이트에서 부여한 점수	-1	없음	+2	-1	없음

3 영화 리뷰에서 '줄리다'라는 단어의 감성 점수는 몇 점을 부여하면 좋을지 생각해 보자. 또, '줄리다'라는 단어가 긍정적인 점수를 받을 수 있는 분야 및 상황에 대하여 찾아보자. (감성 점수는 -2점에서 +2점 범위에서 부여)

영화 리뷰에서 '줄리다'라는 건 영화가 재미없다는 의미이므로 -2점이 적당하다고 생각한다.
또, 침대 상품 리뷰의 경우 '줄리다'라는 표현은 굉장히 긍정적인 의미이므로 +2점을 부여하는 것이 좋다고 생각한다.

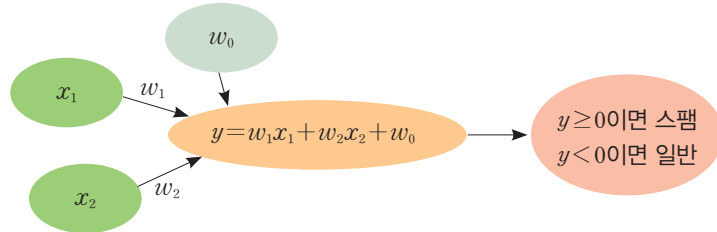
4 일반적인 감성 사전을 만드는 것에 대한 어려움에 대하여 이야기해 보자.

- 2에서 알 수 있듯이 어떤 단어에 얼마의 감성 점수를 부여할 것인지는 주관적인 느낌이 다르기 때문에 정하기가 쉽지 않다.
- 3에서 알 수 있듯이 단어가 모든 상황에서 긍정적이거나 부정적인 감성을 나타내지 않는다는 것을 알 수 있다. 따라서 일반적인 상황에 적용 가능한 감성 사전을 구축하기보다는 분야나 상황에 맞는 감성 사전을 구축할 필요가 있다.



다음 절차에 따라 스팸 메일을 분류해 보자.

하나의 메일에서 스팸 메일에 자주 등장하는 단어 ‘공짜’와 일반 문서에 자주 등장하는 단어 ‘요청’이 나타난 횟수를 각각 x_1, x_2 라고 하자. $y = w_1x_1 + w_2x_2 + w_0$ 일 때, $y < 0$ 이면 일반 메일로 분류하고, $y \geq 0$ 이면 스팸 메일로 분류하려고 한다.



1 4개의 메일에서 확인한 결과가 다음과 같을 때, 이를 만족하도록 하는 w_1, w_2, w_0 의 값을 하나 정하고 친구들과 자신이 구한 값을 비교해 보자. $w_1=1, w_2=-1, w_0=0$

스팸	일반
(5, 3)	(1, 2)
(7, 6)	(3, 5)

2 추가로 4개의 메일에서 확인한 결과가 다음과 같을 때, 이를 만족하도록 w_1, w_2, w_0 의 값을 수정하고 친구들과 자신이 구한 값을 비교해 보자. $w_1=2, w_2=-1, w_0=-2.5$

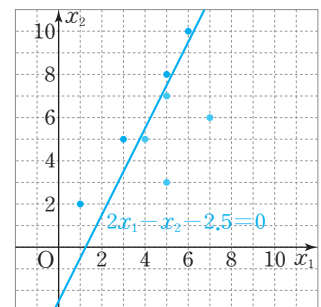
스팸	일반
(4, 5)	(6, 10)
(5, 7)	(5, 8)

• 활동팁 1, 2에서 구한 값은 서로 다를 수 있음을 이해하도록 하고, 효과적으로 분류하도록 w_i 의 값을 정하는 것이 목적임을 이해하도록 한다.

3 2에서 구한 w_1, w_2, w_0 의 값으로 다음 (x_1, x_2) 값에 대하여 메일을 분류해 보자.

	스팸 or 일반
(9, 16)	일반
(7, 11)	스팸
(11, 20)	일반

4 1, 2의 (x_1, x_2) 를 좌표평면 위에 나타내고, 2에서 구한 w_1, w_2, w_0 의 값을 이용하여 직선 $w_1x_1 + w_2x_2 + w_0 = 0$ 을 좌표평면 위에 그려 보자.



5 4에서 구한 직선이 실제 메일을 효과적으로 분류한다고 할 수 있는지 이야기해 보자.

1에서 구한 w_1, w_2, w_0 의 값이 2에서 맞지 않았음을 확인할 수 있듯이 새로운 자료가 추가되면 2에서 구한 w_1, w_2, w_0 의 값도 정확하게 분류하는 값이라고 단정 지을 수 없다. 따라서 많은 자료를 통해 w_1, w_2, w_0 의 값을 정해야 하며 정확하게 직선으로 분류가 불가능할 수도 있다.

• **활동팁** 추가적인 메일 자료를 얻는 경우 하나의 직선으로 정확하게 분류되지 않는 경우가 발생할 수 있으며 이때는 자료와 직선의 차이가 최소가 되는 가중치(w_i)를 찾아야 한다.

시도의 주안점

- 교사는 자료의 유사도를 비교하는 활동을 통해 유사도의 의미를 이해하도록 도움을 준다. 또한 유사도를 계산하는 경우 컴퓨터 프로그램 등을 적극적으로 활용하도록 한다.

핵심 발문

- 유사도를 이용하여 문서를 분류하는 방법은 무엇일까?
- 추천하기 위해서 유사도를 활용할 수 있는 방법은 무엇일까?
- 문서를 분류하기 위해 감성 사전을 어떻게 이용할 수 있을까?
- 문서를 분류하는 알고리즘을 이해할 수 있을까?

교과서 연계 방안

- 교과서를 통해 벡터, 내적, 유사도의 계산 방법 등을 충분히 이해한 후 활동할 수 있도록 한다.



관련 교과 역량

- 의사소통
- 태도 및 실천

- 텍스트 자료의 유사도를 측정하는 방법을 이해할 수 있는가?
- 자료 사이의 유사도를 구하고 비교하는 활동에 어려움을 겪고 있는 모둠원을 배려하고 존중할 수 있는가?

평가 결과 피드백

- 벡터로 표현된 자료의 유사도를 계산하는 방법의 의미를 정확하게 이해하고 있는지 확인하기 위해 설명하거나 토론하는 활동을 하는 기회를 제공한다.
- 유사도를 이용하여 문서 사이의 유사도를 측정하는 과정에서 어떤 수학적 원리를 이용했는지 이해하고 있는지 확인한다.

3. 모듈⑤

이미지 자료의 분류는 어떻게 할까?

가 모듈 개관

성취기준	12인수03-02 인공지능을 이용하여 이미지를 분류하는 수학적 방법을 이해한다.
모듈 설계 의도	성취기준의 의미 인공지능을 이용하여 자료를 분류하는 수학적 방법을 이해한다. 이미지 자료를 분류하는 수학적 방법을 이해한다.
	학습 활동 계획 이 모듈에서는 이미지 자료를 분류하는 수학적 방법을 이해할 수 있도록 필기체 손글씨를 분류하는 방법을 이해할 수 있는 탐구활동을 구성하였다.
	기대하는 학습 결과 모듈별 탐구 학습을 통해 인공지능이 이미지 자료를 분류하는 방법을 이해할 수 있다. 더 나아가 인공지능에서의 수학의 활용과 중요성을 이해하면서 수학에 대한 가치를 알고 흥미를 가질 수 있다.
	모듈 자료에 대한 유의점 인공지능이 이미지 자료를 분류하는 수학적 방법을 익히도록 구성 되어 있으며, 수학적 개념의 계산보다 이해와 적용에 초점을 두고 있다.
교과 역량	의사소통, 문제 해결, 정보 처리, 태도 및 실천
학습 요소	행렬
교수·학습 방법	탐구 학습
교수·학습 방법 및 유의 사항	• 이미지 판별에서는 개와 고양이의 사진을 구별하는 문제 등을 다룰 수 있고, 구체적인 이미지 분류 문제의 해결을 위해 인공신경망과 같은 방법이 개발되었음을 인식하게 한다. • 인공지능과 관련된 수학에 대한 평가를 할 때는 학생들의 흥미, 자신감, 수학의 가치 인식, 학습 태도 등 정의적 영역의 평가와 과정 중심 평가를 할 수 있다.
평가 방법	포트폴리오 평가
평가 방법 및 유의 사항	• 인공지능에 수학이 어떻게 활용되는지 이해하는 것에 초점을 두어 평가한다.

차시	주제	교수·학습 활동	평가 활동
1~2	이미지 자료의 분류	탐구 학습 이미지 자료의 분류 - 숫자 0, 1, 2 분류하기 ○와 × 이미지 분류하기 - MNIST 이해하기 - 이미지 분류 과정에서 행렬의 곱셈이 활용되는 사례 이해하기	포트폴리오 평가 - 이미지를 잘 분류하였는가? - 이미지 분류 과정에서 행렬의 곱셈이 활용되는 사례를 이해하였는가?

1~2차시 | 이미지 자료의 분류

학습 목표	인공지능이 이미지 자료를 분류하는 방법에 대해 이해한다.
수업의 의도	수업 의도 숫자를 나타낸 다양한 그림을 같은 숫자끼리 분류할 수 있도록 하고, 이미지를 구분하는 특징을 발견할 수 있으며, 이미지 분류에 활용되는 행렬의 곱셈을 이해할 수 있도록 설계하였다. 수업의 주안점 실제로 이미지를 분류하는 알고리즘을 정확하게 이해하는 것은 쉽지 않으므로 여기서는 단계를 간단히 하여 제시하였고, 학생들이 직관적으로 이해할 수 있도록 한다. 수업 활동 시 유의점 이미지를 분류할 때 활용되는 과정을 그대로 적용시키기 위해서는 파이썬을 활용할 수 있으나, 파이썬을 활용하기 쉽지 않으므로 간단한 활동을 통해 인공지능으로 어떻게 구현할 수 있는지 이해할 수 있도록 한다. 평가의 주안점 복잡한 계산은 컴퓨터를 활용할 수 있도록 하고, 자신이 이해한 바를 설명할 수 있는지에 초점을 두어 평가한다.
준비물	활동지, 스마트폰(태블릿 PC로 대체 가능)

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	• 4~5인 1개 모둠 구성	
수행	모둠 활동1	- 보기의 그림들을 0, 1, 2로 분류하기 ○를 나타낸 이미지와 ×를 나타낸 이미지의 모양을 구분하는 특징 찾기	포트폴리오 평가
	모둠 활동2	- MNIST 알아보기 - 손글씨로 쓴 숫자를 컴퓨터가 인식하는 방법에 대해 이해하기	
	모둠 활동3	이미지 분류 과정에 행렬의 곱셈이 활용되는 사례 이해하기	
정리		이미지 분류 과정에 행렬의 곱셈이 활용됨을 이해	



이미지 자료의 분류

년

월

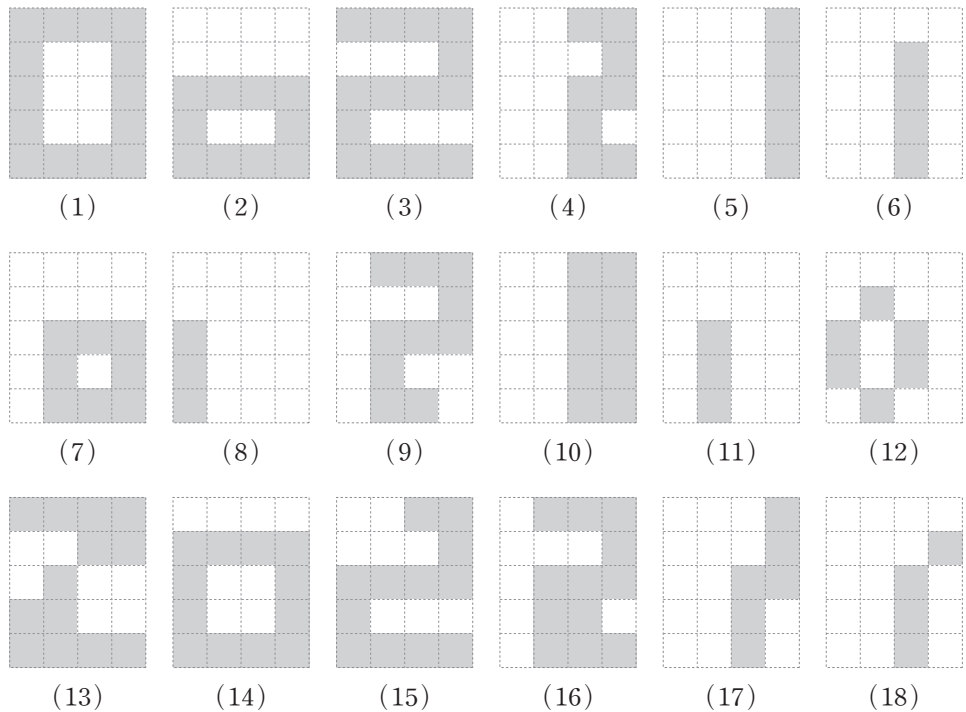
일



이미지를 분류하는 방법을 알아보자.

디지털 카메라의 얼굴 인식, 출입문의 지문 인식, 전자상거래 사이트의 상품 인식, 자율주행 자동차의 도로 상황 파악 등 이미지를 판별하고 분류하는 인공지능 기술은 일상생활의 다양한 곳에서 사용되고 있다.

1 주어진 그림들이 숫자 0, 1, 2 중 어떤 숫자를 나타낸 것인지 분류하여 보자.



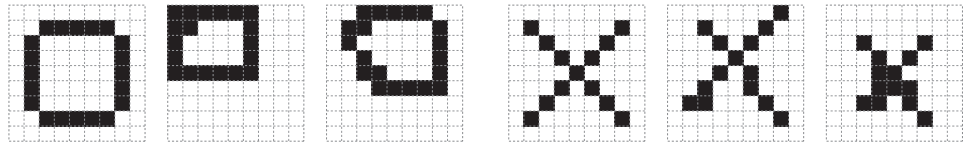
개인이차기 있을 수 있으나 0으로 분류할 수 있는 것은 (1), (2), (7), (12), (14)

1로 분류할 수 있는 것은 (5), (6), (8), (10), (11), (17), (18)

2로 분류할 수 있는 것은 (3), (4), (9), (13), (15), (16)이다.

(17)과 (18)은 0, 1, 2 중 어느 것에도 속하지 않는다고 답할 수도 있으나 이 셋 중 하나로 반드시 분류해야 한다면 1로 분류할 수 있다.

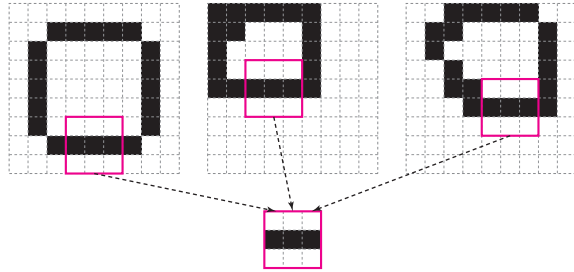
2 다음은 ○를 나타낸 이미지와 ×를 나타낸 이미지를 나타낸 것이다. 이 두 가지 모양을 구분하는 특징을 찾아보자.



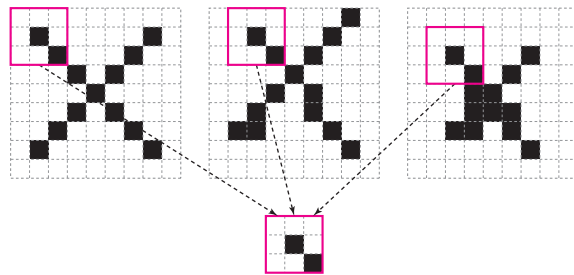
[그림 1] ○를 나타낸 이미지

[그림 2] ×를 나타낸 이미지

○를 나타낸 이미지는 [그림 3]과 같이 평평한 수평선이 있고, ×를 나타낸 이미지는 [그림 4]와 같이 대각선으로 색칠한 격자가 나타난다는 특징이 있다.



[그림 3] ○를 나타낸 이미지의 특징



[그림 4] ×를 나타낸 이미지의 특징

• 활동팁 이미지에 어떤 특징이 있는지 알아내는 방법은 이미지 위에서 필터를 조금씩 움직이면서 모든 화소에 필터를 적용하고 특징을 찾는다.

슬라이딩 윈도우

0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	1	0	0
0	0	1	0	0	0	1	0	0	0
0	0	0	1	0	1	0	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0	0	1	0	1	0	0	0	0
0	0	1	0	0	0	1	0	0	0
0	1	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0

필터

0	0	0
0	1	0
0	0	1

필터링 결과

$$\begin{aligned}
 &= (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (1 \times 1) \\
 &\quad + (0 \times 0) + (0 \times 0) + (0 \times 0) + (1 \times 1) \\
 &= 2
 \end{aligned}$$

슬라이딩 윈도우

0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	1	0	0
0	0	1	0	0	0	1	0	0	0
0	0	0	1	0	1	0	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0	0	1	0	1	0	0	0	0
0	0	1	0	0	0	1	0	0	0
0	1	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0

필터

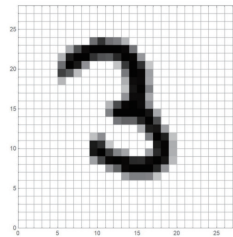
0	0	0
0	1	0
0	0	1

필터링 결과

$$\begin{aligned}
 &= (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 1) \\
 &\quad + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 1) \\
 &= 0
 \end{aligned}$$

MNIST⁵⁾를 알아보자.

MNIST(Modified National Institute of Standards and Technology database)는 손으로 쓴 숫자들로 이루어진 데이터베이스로, 다양한 화상 처리 시스템을 트레이닝하기 위해 널리 사용된다. ‘MNIST 데이터 셋’의 숫자 이미지는 가로×세로의 크기가 28(픽셀)×28(픽셀)로 이루어져 있어 픽셀의 개수는 총 784개이다. 각 픽셀은 $x_0, x_1, x_2, \dots, x_{783}$ 과 같이 표현하면 $x_i=0$ 또는 $x_i=1$ 로 표현할 수 있다. 이때, $x_i=1$ 인 픽셀들의 위치가 해당 숫자가 그려진 위치가 된다.



MNIST 손글씨 예시

손글씨의 인식 과정을 개략적으로 다음과 같은 단계로 이루어진다.

- ① 60,000개의 학습 데이터들을 0부터 9까지의 10개의 그룹으로 분류한 다음, 각 그룹에 속한 이미지들에서 픽셀에 입력된 채널값의 분포를 계산하여 채널값이 주로 낮은 픽셀의 위치에 가중치를 부여하게 된다.

0그룹의 픽셀별 가중치 : $w_{(0,0)}, w_{(0,1)}, w_{(0,2)}, \dots, w_{(0,783)}$

1그룹의 픽셀별 가중치 : $w_{(1,0)}, w_{(1,1)}, w_{(1,2)}, \dots, w_{(1,783)}$

⋮

9그룹의 픽셀별 가중치 : $w_{(9,0)}, w_{(9,1)}, w_{(9,2)}, \dots, w_{(9,783)}$

- ② 입력된 손글씨 이미지의 각 픽셀의 채널값을 $x_0, x_1, x_2, \dots, x_{783}$ 로 나타낸다.
- ③ 픽셀별 가중치와 입력된 이미지의 해당 픽셀의 채널값을 곱한 후 모두 더하여 각 숫자별로 인식될 가능성을 계산한다.

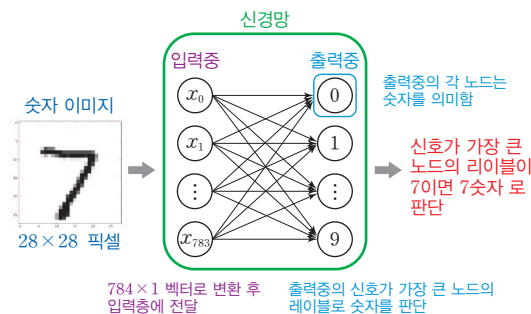
0으로 인식될 가능성 : $w_{(0,0)}x_0 + w_{(0,1)}x_1 + w_{(0,2)}x_2 + \dots + w_{(0,783)}x_{783}$

1로 인식될 가능성 : $w_{(1,0)}x_0 + w_{(1,1)}x_1 + w_{(1,2)}x_2 + \dots + w_{(1,783)}x_{783}$

⋮

9로 인식될 가능성 : $w_{(9,0)}x_0 + w_{(9,1)}x_1 + w_{(9,2)}x_2 + \dots + w_{(9,783)}x_{783}$

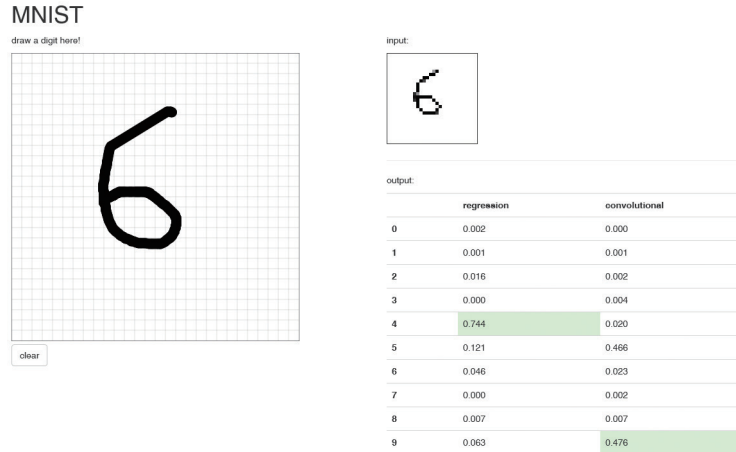
- ④ 가능성이 가장 높은 값의 숫자로 인식한다.



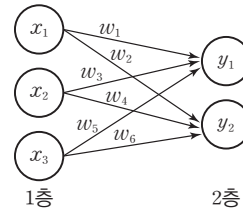
5) <http://naver.me/5w54NGG2>

1 사이트(<https://tensorflow-mnist.herokuapp.com>)에 접속하여 숫자를 써 보고 컴퓨터가 어떻게 인식하는지 확인해 보자.

예 숫자를 쓰는 곳에 0~9까지 중 임의의 숫자를 기입한다. [clear]를 누르면 숫자를 삭제하고 다른 숫자를 쓸 수 있다. 사이트(<https://tensorflow-mnist.herokuapp.com>)에 들어가 6을 입력하면 다음과 같이 표현된다.



2 1층에 있는 3개의 인공 뉴런이 출력하는 값이 각각 x_1, x_2, x_3 이고, 가중치가 다음과 같다고 하자.



2층에 있는 2개의 인공 뉴런에 입력되는 값을 각각 y_1, y_2 라 할 때,

$$x_1w_1 + x_2w_3 + x_3w_5 = y_1$$

$$x_1w_2 + x_2w_4 + x_3w_6 = y_2$$

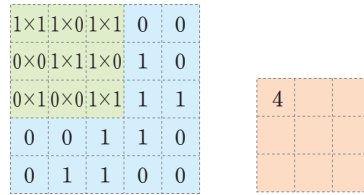
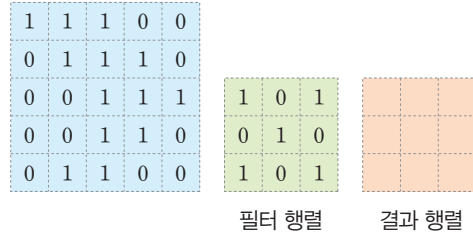
이다. 이를 행렬의 곱셈으로 나타내면 다음과 같다.

$$\begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \\ w_5 & w_6 \end{pmatrix} = \begin{pmatrix} y_1 & y_2 \end{pmatrix}$$

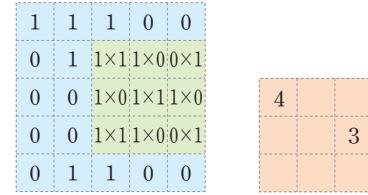
이를 이용하여 출력되는 값을 $y_0, y_1, y_2, \dots, y_9$ 라고 할 때, 행렬로 표현해 보자.

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_9 \end{pmatrix} = \begin{pmatrix} w_{(0,0)} & w_{(0,1)} & \cdots & w_{(0,9)} \\ w_{(1,0)} & w_{(1,1)} & \cdots & w_{(1,9)} \\ w_{(2,0)} & w_{(2,1)} & \cdots & w_{(2,9)} \\ w_{(3,0)} & w_{(3,1)} & \cdots & w_{(3,9)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{(9,0)} & w_{(9,1)} & \cdots & w_{(9,9)} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_9 \end{pmatrix}$$

3 5×5 행렬은 9개의 서로 다른 3×3 행렬로 분해할 수 있다. 다음과 같이 5×5 행렬에서 9개의 서로 다른 3×3 행렬에 3×3 필터 행렬을 곱한 모든 성분의 합을 마지막 결과 행렬에 나타내 보자. 예를 들어 [그림 1]과 같이 5×5 행렬의 첫 번째 3×3 행렬과 3×3 필터 행렬을 곱한 값의 성분의 합은 4이므로 결과 행렬의 (1, 1)에 4를 쓰고, [그림 2]와 같이 5×5 행렬의 두 번째 줄의 맨 끝의 3×3 행렬과 3×3 필터 행렬을 곱한 값의 성분의 합은 3이므로 결과 행렬의 (2, 3)에 3을 쓴다.

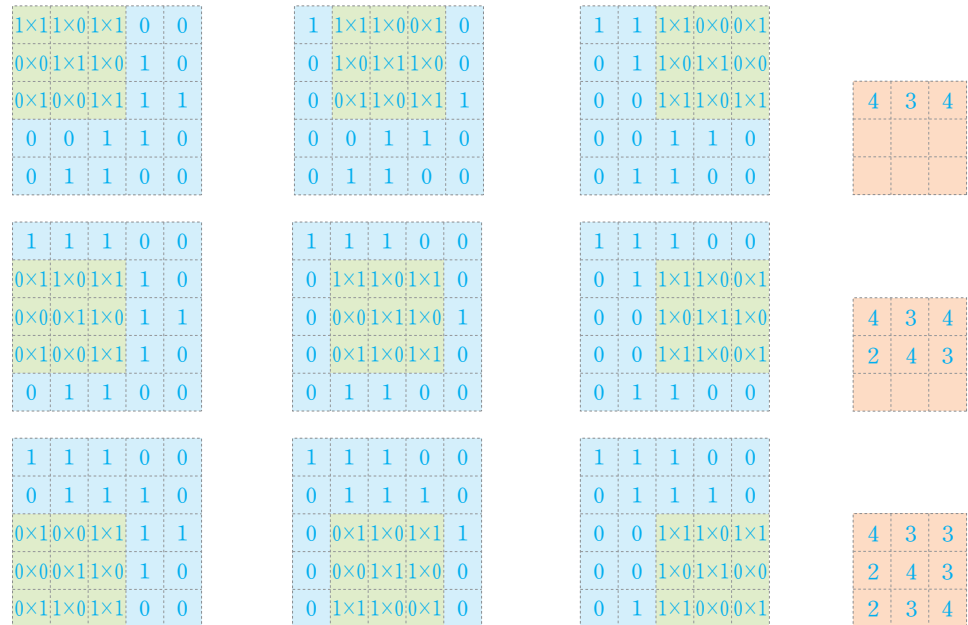


[그림 1]



[그림 2]

계산 결과는 다음과 같다.



시도의 주안점

- 인공지능이 이미지 자료를 분류하는 방법에 행렬의 곱셈이 활용됨을 이해할 수 있도록 한다.
- 활동지에는 행렬의 곱셈에 대한 내용을 제시하지 않았으므로 이 부분은 별도로 학습시킨다.

핵심 발문

- 이미지를 분류하는 과정에 활용되는 수학적 개념은 무엇일까?
- 행렬의 곱셈이 이미지 자료를 분류하는 데 어떻게 활용될까?

교과서 연계 방안

- 활동 후에 다양한 경우에서 활용되는 행렬의 곱셈과 관련된 문제를 해결하고 탐구할 수 있도록 한다.



관련 교과 역량

- 의사소통
 - 문제를 어떻게 해결하였는지 그 이유를 논리적으로 설명할 수 있는가?
- 정보 처리
 - 공학적 도구를 활용하여 자료를 목적에 맞게 사용할 수 있는가?

평가 결과 피드백

- 그림을 분류하는 과정에서 활용되는 행렬의 곱셈을 제대로 계산할 수 있는지 확인하고, 잘못 계산한 경우 학생이 스스로 오류를 찾고 수정할 수 있게 한다.
- 그림을 분류하고 그 이유를 설명하는 과정에서 자신의 답안의 근거를 효과적으로 발표한 학생의 활동 사례를 학교생활기록부에 기록한다.

4. 모듈⑥

확률을 이용한 예측은 어떻게 할까?

가 모듈 개관

성취기준	12인수03-03 자료를 분석하여 사건이 일어날 확률을 구하고 예측에 이용할 수 있다.
모듈 설계 의도	성취기준의 의미 자료를 분석하여 사건이 일어날 통계적 확률을 구하고 자료에 나타난 여러 조건에 따른 확률을 비교하여 다양한 예측 활동을 할 수 있다. 자료로부터 확률을 구하는 활동을 통해 인공지능이 수행하는 예측에 확률이 사용됨을 이해하게 한다. 학습 활동 계획 이 모듈은 인공지능의 예측에 확률이 활용됨을 이해하게 하며, 특히 자료로부터 구하는 통계적 확률의 의미를 이해하는 것을 목표로 한다. 먼저 틱택토 인공지능, 알파고 등을 사례로 활용하여 인공지능이 확률 시뮬레이션을 활용하는 원리에 대해 기초적인 수준에서 이해하고 체험해 보게 한다. 유튜브 인기 동영상 자료 등을 이용하여 자료에서 여러 조건에 따른 통계적 확률을 구해 보고, 구한 확률을 이용하여 예측, 판단하는 활동을 하게 한다. 조건부확률을 이용한 스팸 메일 필터의 원리를 비형식적으로 이해할 수 있도록 하였다. 기대하는 학습 결과 인공지능의 예측에 확률이 이용됨을 이해한다. 자료를 이용하여 사건의 확률을 구하고 해석할 수 있다. 모듈 자료에 대한 유의점 공학적 도구를 활용하므로 컴퓨터나 태블릿PC 등이 사용 가능한 환경에서 수업을 진행해야 한다.
교과 역량	추론, 문제 해결, 의사소통, 정보 처리
학습 요소	확률
교수·학습 방법	매체 및 도구 활용 학습, 탐구 학습
교수·학습 방법 및 유의 사항	확률은 가능성을 예측하고 싶은 사건을 설정하고 자료를 수집하여 그 사건이 일어날 확률을 구하며, 조건부확률의 용어와 기호를 도입하지 않고 특정 조건으로 세분화된 확률을 상대도수로 추정하고 예측하는 과정에 어떤 수학적 원리가 사용되는지 경험하는 정도로 간단히 다룬다.

평가 방법		포트폴리오 평가	
평가 방법 및 유의 사항		• 자료의 예측에 대한 평가는 확률 계산이나 관계식 등 수치 계산에 치중하지 않도록 하고, 자료의 경향성을 추세선으로 나타내는 활동과 확률을 예측에 이용하는 과정에 중점을 둔다.	
차시	주제	교수·학습 활동	평가 활동
1	자료와 확률	매체 및 도구 활용 학습 확률 시뮬레이션을 이용하여 예측하기, 실생활 자료로부터 통계적 확률 구하기 • 컴퓨터와 텍토티 게임해 보기 • 시뮬레이션 결과를 해석하여 컴퓨터의 입장에서 텍토티의 다음 수를 결정해 보기 • 빅데이터 유튜브 자료 살펴보기 • 유튜브 자료를 분석하여 예측과 의사 결정에 활용하기 • 인공지능에서 언어 모델의 활용에 대해 알아보기	포트폴리오 평가 • 시뮬레이션 결과를 해석하고, 이를 예측에 활용할 수 있는가? • 자료로부터 구한 통계적 확률을 이용하여 의사 결정을 내리고 그 이유를 논리적으로 설명할 수 있는가? • 공학적 도구를 활용하여 자료를 목적에 맞게 가공할 수 있는가?
2	스팸 메일 필터	탐구 학습 스팸 메일 필터의 이해 • 스팸 메일과 햄 메일 이해하기 • 스팸 메일 필터의 학습과 분류 과정 이해하기 • 특정 자료를 이용하여 학습한 스팸 메일 필터의 스팸 판정 관련 문제 해결하기 • 2개의 단어를 이용한 스팸 판정에 관한 문제 해결하기	포트폴리오 평가 • 스팸 메일 필터의 학습과 예측 과정을 이해하는가? • 특정 단어 1개가 제목에 포함된 메일이 스팸 메일일 확률을 구할 수 있는가? • 특정 단어 2개가 제목에 포함된 메일이 스팸 메일일 확률을 구할 수 있는가?

나 차시별 교수·학습 자료

1 차시 | 자료와 확률

학습 목표	자료로부터 여러 가지 사건의 확률을 구하고 그 의미를 이해한다.
수업의 의도	수업 의도 틱택토 사례를 이용하여 알파고에 활용된 몬테카를로 트리 검색 기법을 이해할 수 있도록 하였다. 또 학생들에게 친숙한 유튜브 자료를 이용하여 여러 가지 사건의 통계적 확률을 구하고 예측에 활용할 수 있게 하였다. 특히, 사건의 조건에 따라 자료의 일부만을 전체 경우로 다루는 문제를 해결하면서 비형식적으로 조건부확률을 이해할 수 있게 설계하였다. 수업의 주안점 자료로부터 확률을 구하고 예측 활동 및 비형식적인 통계적 추론을 수행할 수 있게 한다. 수업 활동 시 유의점 컴퓨터나 태블릿PC를 이용할 수 있게 한다. 평가의 주안점 복잡한 계산은 컴퓨터를 활용할 수 있게 하고, 논리적 과정을 중심으로 평가한다.
준비물	활동지, 노트북(태블릿 PC나 스마트폰으로 대체 가능)

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	• 2~4명을 1개 모둠으로 구성 • 활동지 배부	
수행	모둠 활동1	• 컴퓨터와 틱택토 게임하기 • 시뮬레이션 결과를 해석하여 컴퓨터의 입장에서 틱택토의 다음 수를 결정하기 • 알파고에 사용된 확률 시뮬레이션 기법 이해하기	포트폴리오 평가
	모둠 활동2	• 유튜브 자료를 분석하여 예측과 의사 결정에 활용하기 • 인공지능에서 언어 모델의 활용에 대해 알아보기	
정리		• 내용 정리	



자료와 확률

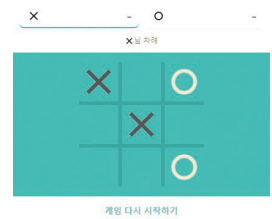
년 월 일



컴퓨터와 틱택토(tic-tac-toe) 게임을 해 보자.

틱택토(tic-tac-toe)는 두 명이 번갈아 가며 ○와 ×를 3×3 판에 써서 먼저 같은 글자를 가로, 세로, 혹은 대각선 상에 먼저 놓는 사람이 이기는 게임이다.

Google 검색엔진에서 '틱택토'를 검색하거나 사이트(<https://www.google.com/search?q=tic+tac+toe>)에 접속하면, 인공지능인 컴퓨터와 틱택토 대결을 할 수 있다.



1 컴퓨터와 틱택토 대결을 3판 해 보자. 자유롭게 컴퓨터와 틱택토를 하면서 규칙에 익숙해질 수 있도록 한다.

컴퓨터는 틱택토에서 다음 위치를 결정할 때, 각 위치마다 무작위 시뮬레이션을 실행해 본 후 승리할 확률이 가장 높은 위치를 찾는다. 이러한 방식을 몬테카를로 트리 검색이라고 한다. 예를 들어 다음 상황을 보자.

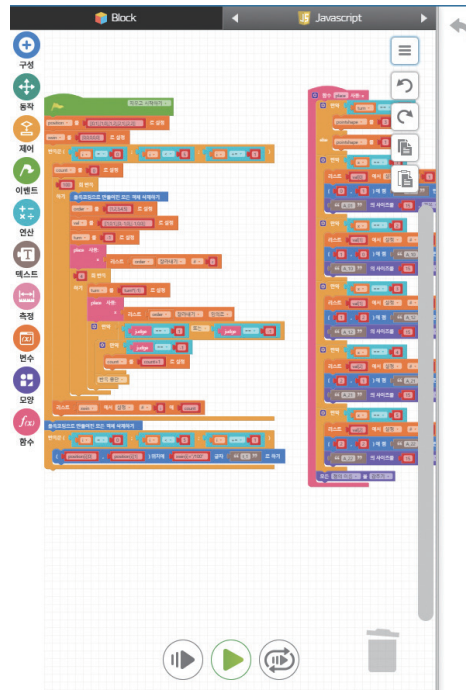
×가 다음 수를 둘 때 가장 유리한 위치는?

●	③	⑤
①	×	④
●	②	×

×(컴퓨터)가 다음 수를 둘 수 있는 위치는 ①, ②, ③, ④, ⑤로 총 5개이다. 컴퓨터는 ①에 ×를 놓은 후, 나머지 4번의 수를 무작위로 두는 시뮬레이션을 100번 하여 승리한 횟수를 기록한다. 마찬가지로 ②~⑤에 대해서도 100번씩 동일한 방식으로 시뮬레이션을 하여 승리한 횟수를 기록한 후, 승리할 확률이 가장 높은 곳을 선택한다.



2 (AI 웹실현실)의 '몬테카를로 시뮬레이션: 틱택토'에서 ▶ 버튼을 눌러 블록 코딩을 실행하면 위에서 설명한 시뮬레이션을 실행해 볼 수 있다. 시뮬레이션 결과를 적은 후 컴퓨터의 입장에서 어떤 위치를 선택하는 것이 유리한지 적어 보자.



X가 다음 수를 둘 때 가장 유리한 위치는?

●		
	×	
●		×

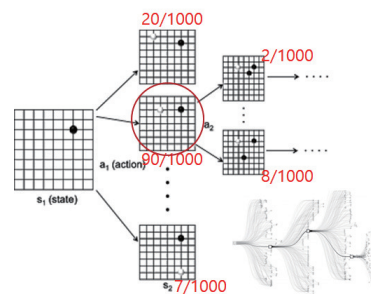
예 예를 들어, 시뮬레이션 결과가 다음과 같이 나왔다면, 가장 왼쪽 위치를 선택하는 것이 컴퓨터에게 유리한 수라고 판단할 수 있다. 실제로도 가장 맨 왼쪽 위치를 × 차례에서 막지 않으면 다음 차례에서 ○가 이기게 된다.

●	52/100	26/100
63/100	×	45/100
●	23/100	×

[시뮬레이션 결과]

알파고와 확률을 알아보자.

19×19 바둑판에서 바둑의 경우의 수는 우주의 원자 개수보다 많다고 한다. 따라서 컴퓨터의 성능이 아무리 좋다고 하여도 모든 경우의 수를 따져가며 바둑을 둘 수 없다. 알파고에서도 앞의 틱택토에서 설명한 확률 시뮬레이션 기법이 활용되었다. 알파고는 다음 수를 결정할 때, 각 위치마다 무작위 시뮬레이션을 1000번 실행하여 승리



한 횟수를 구한다. 위치별로 승리할 확률 $\left(\frac{\text{승리한 횟수}}{1000\text{번}}\right)$ 을 계산하여 확률이 가장 높은 위치에 다음 수를 둔다. 이와 같은 방법은 알파고 이전에 이미 바둑 인공지능에서 널리 활용되는 것이었지만, 알파고는 이 확률 기법에 딥러닝을 추가로 활용하여 실력을 높였고, 결국 사람을 이기게 되었다.

1 틱택토나 알파고 알고리즘에서 활용된 확률은 '주사위를 던졌을 때 2가 나올 확률 $\frac{1}{6}$ '과 어떤 차이가 있는가?

예 주사위 확률은 던져보지 않아도 알 수 있지만 바둑이나 틱택토에서 활용된 확률은 시뮬레이션을 통해 자료를 생성한 후 상대도수로 구한 것이라는 데 차이가 있다.

현실 세계에서 주사위와 동전을 던질 때의 확률과 같이 실험을 해보지 않고도 확률을 구할 수 있는 경우는 드물다. 중학교 때 시행 횟수가 충분히 커지면 어떤 사건이 일어나는 상대도수가 이론적으로 구한 확률의 값과 비슷해진다는 것을 배웠다. 따라서 이론적으로 확률을 구할 수 없는 현실 세계의 대부분의 사건은 누적된 자료를 이용하여 상대도수를 구하여 가능성을 나타내는데, 이 값을 경험적 확률 또는 통계적 확률이라고 한다.

지수의 꿈은 유튜버이다. 그는 어떤 카테고리에서 어떤 콘텐츠의 영상을 올려야 인기 동영상 만들 수 있는지 알아보기 위하여 최근 구글에서 공개한 유튜브(Youtube) 인기 동영상 빅데이터⁶⁾를 분석하기로 하였다.

사이트(<https://b.ly/2WjzZ24>)에 접속하면 다음과 같이 빅데이터를 확인할 수 있다.

6) 본 차시에서 사용되는 데이터들은 <https://www.kaggle.com/datasnaek/youtube-new>의 일부를 학습에 적합하게 가공한 것이다

YouTubeRawDataCSV																			🔍	👤
파일 열기 100% 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245																				

만약 우리가 지수라면, 이 유튜브 데이터를 통해 무엇을 확인해 보고 싶을까? 자유롭게 아이디어를 이야기해 보자.

예 조회수가 가장 많은 카테고리를 확인해 볼 것이다.

조회수 대비 좋아요 수가 높은 영상이나 댓글이 많이 달리는 카테고리를 확인할 것이다.



엑셀을 이용하여 확률을 구하는 방법을 알아보자.

사이트(<https://bit.ly/3mUnMwN>)에 접속하면 다운로드할 수 있는 데이터는 구글에서 제공한 원자료 중에 5개의 카테고리(동물, 교육, 과학기술, 스타일, 게임)별로 10개씩 무작위로 선택한 것이다. 단, 지수는 개인 유튜버이기 때문에 방송사나 대기업에서 올린 광고 영상 등은 제외하고 개인 유튜버들의 영상 중에서만 선별하였다. 구글 스프레드시트나 엑셀로 파일을 열면 아래와 같은 화면을 볼 수 있다.

A	B	C	D	E	F	G	H	I
채널	제목	카테고리	태그	업로드일	조회수	좋아요	싫어요	댓글수
동물들의사랑	아기 기린 한 마리	동물	동물들의사랑	1	140651	221	48	65
Ani는 고양이 내가 주인	아리(영양) 아리(영양) 더 비기닝 The Sheriff AF	동물	아리(영양) 아리(영양) 더 비기닝 The Sheriff AF	1	55124	4138	9	642
haha ha	카라와 같이	동물	카라와 같이	1	74847	3462	18	710
FeelSoGood	감자지개 주인이 사라지는 마술을 보여줬다!	동물	감자지개 주인이 사라지는 마술을 보여줬다!	1	219399	3299	54	275
소녀의형제GirlsPlanet	서로의 간식을 탐내는 감자지개	동물	감자지개 소녀의형제GirlsPlanet	2	206222	5245	51	1022
관찰남	새벽5시만 되면 방송연습하는 고양이	동물	고양이 관찰남	2	43516	1811	21	394
개감들이만	마리의 방송을 보는 작은 감자지개	동물	개감들이만	3	30833	327	13	25
수리노출RunNoel	간식 달라고 아우성치는 귀순 고양이	동물	수리노출RunNoel	3	59535	2695	13	691
Bodeum official	[강행록의 소소한 Q&A] 당신도 호아인인가요?	동물	강행록의 소소한 Q&A	1	28625	1192	4	659
정브로	한국에서 열대야가 진짜 강하다고9국내최고 1000 동물	동물	정브로	1	101657	1287	49	793
복숭아 워터주는 남자	[정해리 Live]복숭아 워터주는 남자	동물	정해리	2	21497	874	42	33
늘이이디이 오리	장고로 연주하는 귀순 고양이	동물	늘이이디이 오리	3	1407497	10055	2127	5937
김정일국립현대미술관	김정일의 자서전 작품관 2019년 05월 1일	예술	김정일의 자서전 작품관 2019년 05월 1일	2	39743	3053	87	454
코코로코	왜 아홉 무렵엔 푸른색인지 물었다	동물	코코로코	2	541304	6096	181	738
꽃노점	같이 공부해요study with me(정지현)정지현의소고	음악	같이 공부해요study with me(정지현)정지현의소고	3	179943	2653	58	342
홍문	한국 문화에 관심이 많은 조지현	음악	홍문	2	54295	217	15	48
제비시	유명한 사람이 유머러스하게	음악	제비시	4	61232	1378	59	87
디바제시DeevaJessica	[토미] 3000명의 영국군인들이 유령처럼 사라졌다	음악	디바제시DeevaJessica	2	82349	1489	56	293
Charity Baptist Church	트럼프 이전 백악관 파기와 성경 예언에 따른	종교	트럼프 이전 백악관 파기와 성경 예언에 따른	1	20153	0	0	68

구글 스프레드시트나 엑셀 등의 스프레드시트 프로그램에서 다음과 같은 방법으로 특정 영역 셀들의 합을 구할 수 있다.

SUM					=sum(B4:B8)
A	B	C	D	E	
1					
2					
3					
4		1			
5		2			
6		3			
7		4			
8		5			
9		=sum(B4:B8)			
10					



- 1 지수가 유튜브를 시작한다고 했을 때, 어떤 카테고리에서 시작하는 것이 유리하다고 생각되는가? 이유를 함께 제시해 보자.

스프레드시트에서 sum 함수를 이용해 카테고리별 10개 영상의 조회수를 모두 더해 보면 다음과 같다.

	조회수
동물	959,779
교육	2,487,612
과학기술	2,019,572
스타일	1,282,929
게임	1,603,070

따라서 지수는 교육 또는 과학기술 카테고리에 해당하는 콘텐츠로 유튜브를 시작하는 것이 유리할 것이다.



- 2 지수가 시험 삼아 유튜브에 동영상 하나 올렸다. 지수의 영상을 본 시청자가 좋아요를 누를 확률을 주어진 데이터와 스프레드시트를 이용하여 구해 보자.

50개 영상의 좋아요 총합을 조회수 총합으로 나누면 $\frac{152948}{8352962} = 0.0183\cdots$ 이므로 약 1.83%이다.



- 3 영상을 본 시청자에게 좋아요를 받기 유리하려면 지수는 어떤 카테고리에 영상을 업로드하는 것이 유리할까? 스프레드시트를 이용하여 필요한 확률들을 구하고 그것을 근거로 설명해 보자.

영상을 좋아한 사람이 좋아요를 누를 확률을 카테고리별로 구하면 다음과 같다.

	조회수	좋아요	$\frac{\text{좋아요}}{\text{조회수}}$
동물	959,779	24,957	0.02600286108
교육	2,487,612	28,831	0.01158982993
과학기술	2,019,572	19,497	0.009654025704
스타일	1,282,929	48,666	0.03793350996
게임	1,603,070	30,997	0.019336024

따라서, 스타일, 동물 카테고리 순으로 좋아요를 받기 유리하다는 것을 알 수 있다.

- 4 2와 3은 모두 좋아요를 받을 확률과 관계된다. 2와 3에서 구한 확률의 차이점에 대해 친구들과 토론해 보자.

2에서 구한 확률은 전체 조회수 총합에서 전체 좋아요 총합의 비율로 구한 확률이다. 그러나 3의 확률은 좋아요를 받기 유리한 카테고리를 알아내기 위해 카테고리별로 조회수 대비 좋아요의 비율을 구했다는 점이 다르다.



5 지수가 올린 영상이 이틀이 지났지만 인기 동영상에 되지 못했다. 지수가 올린 영상이 인기 동영상이 될 확률은 얼마인가? 여러분이 지수라면 어떻게 대처할 것인지 생각해 보자.

50개의 영상 중에 2일 이후 인기 동영상이 된 영상은 6개 뿐이다. 따라서 인기 동영상이 될 확률은 $\frac{6}{50} = 12\%$ 로 낮으므로 지수의 입장이라면 이전 영상이 인기가 없던 요인을 분석하고 그 결과를 반영하여 새로운 영상을 준비할 것이다.



언어 모델에 대해 알아보자.

검색 엔진에 단어를 입력하다보면 다음과 같이 추천 검색어가 제시되는 것을 볼 수 있다. 이러한 것은 주어진 이전 단어들로부터 다음 단어를 예측하는 언어 모델이라는 것과 관련되어 있다.



언어 모델(Language Model, LM)은 언어라는 현상을 모델링하고자 단어의 순서에 확률을 할당하는 모델이다. 언어 모델을 만드는 방법은 크게는 통계를 이용한 방법과 인공신경망을 이용한 방법으로 구분할 수 있다. 최근에는 통계를 이용한 방법보다는 인공신경망을 이용한 방법이 더 좋은 성능을 보여 주고 있다. 최근 화제가 되고 있는 자연어 처리 신기술인 GPT도 인공신경망 언어 모델의 개념을 사용하여 만들어졌다. 자연어 처리에서 단어 순서에 확률을 할당하는 일은 특히 중요하다. 예를 들어 자동 번역 엔진은 “나는 수학을 배운다.”와 “나는 수학을 배워진다.” 중에 전자의 확률이 더 높으므로 전자를 택한다. 음성 인식에서도 “나는 수학을 좋아한다.”와 “나는 수학을 조아한다.” 중에 전자의 확률이 더 높으므로 전자를 택한다. 언어 모델은 이와 같이 언어와 관련된 인공지능 처리에서 확률을 이용해 보다 적절한 문장을 선택하는 데 활용된다.

1 언어 모델이 다음과 같은 25개의 자료를 이용하여 학습을 하였다고 가정하자.

운동을 시작해야 하는 이유 8가지
올바른 운동으로 새해에도 건강하게
오래 살기 위해 운동을 해야 하는 진짜 이유
당신이 운동을 시작해야 하는 이유
운동을 해야 하는 진짜 이유
헬스장 운동 시 꼭 알아야 할 8가지
운동을 왜 하시나요?
하나만으로도 충분한 당신이 운동을 해야 하는 이유
운동을 통해 체중 조절을 할 수 있습니다.
근육은 젊음 찾아주는 회춘약……. 가장 쉬운 근력운동 방법
운동을 통해 몸짱 만들기
운동을 안하는 이유
우리가 운동을 통해 얻을 수 있는 이점 8가지
운동을 20분씩 매일 꾸준히 하면 두뇌 활동이 극적으로 활발해집니다.
운동을 꾸준히 그리고 즐기면서 하면 몸의 가벼움과 상쾌함을 느끼실 수 있을 것입니다.
운동은 밥이다.
운동을 하는 것이 중요한 이유에는 신체적 건강부터 만족감까지 다양한 부분…….
운동을 하지 않는 이유
운동을 해요
운동이 면역력에 미치는 효과 분석
운동을 습관처럼 반복하세요.
운동을 하려다가 왜 늘 망하는가?
새해가 되면 운동을 결심하고, 건강검진 전후로 반짝 몸에 신경쓴다.
구체적으로 운동을 기록하는 방법은 다음과 같습니다.
나이가 들수록 근력 운동을 해야 하는 이유를 과학적으로 설명해주세요.

이 언어 모델은 “운동을” 다음에 어떤 단어를 추천할 것인가? 순서대로 2가지만 찾고 그 단어가 “운동을” 다음에 나올 확률을 구해 보자.

“운동을”이라는 단어가 포함된 문장은 총 20개이다. 따라서 해야 $\left(\frac{4}{20}\right)$, 통해 $\left(\frac{3}{20}\right)$ 순으로 추천할 것이다.

지도의 주안점

- 빅데이터를 실제로 경험하고, 자료로부터 구한 통계적 확률을 이용하여 의사결정에 활용하는 경험을 하게 한다.
- 사건에 따라서 전체 자료 중 특정 조건을 만족하는 자료만을 이용하여 확률을 구해야 한다는 조건부확률 개념을 비형식적으로 이해하고 구해볼 수 있게 한다.
- 짝, 모둠 활동을 통해 알지오매스, 스프레드시트 등 컴퓨터 프로그램에 익숙하지 않은 학생들도 어려움 없이 활동에 참가할 수 있게 한다.

핵심 발문

- 동전의 앞면이 나올 확률 $\frac{1}{2}$ 과 유튜브 영상이 인기영상이 될 확률은 어떤 차이가 있을까?
- 틱택토 시뮬레이션 결과가 매번 달라지는 이유는 무엇일까?

교과서 연계 방안

- 활동 후에 확률, 조건부확률과 관련된 교과서 내용을 학습할 수 있다.



관련 교과 역량

- 의사소통
 - 자료로부터 구한 통계적 확률을 이용하여 의사 결정을 내리고 그 이유를 논리적으로 설명할 수 있는가?
- 태도 및 실천
 - 공학적 도구를 활용하여 자료를 목적에 맞게 가공할 수 있는가?

평가 결과 피드백

- 주어진 자료나 시뮬레이션 결과로부터 통계적 확률을 수학적으로 계산할 수 있는지 확인하고, 잘못 계산한 경우 학생이 스스로 오류를 찾고 수정할 수 있게 한다.
- 확률을 구하고 의사 결정을 내리는 활동에서, 자신의 답안의 근거를 효과적으로 발표한 학생의 활동 사례를 학교생활기록부에 기록한다.

2차시 | 스팸 메일 필터

학습 목표	확률을 이용한 스팸 메일 필터의 원리를 이해한다.
수업의 의도	수업 의도 간단한 자료를 이용하여 비형식적으로 조건부확률을 계산해보면서 스팸 메일 필터의 학습과 분류 과정을 이해할 수 있도록 구성하였다. 수업의 주안점 작은 자료를 이용하여 직접 특정 단어 1개 또는 단어 2개가 제목에 포함된 메일을 스팸 메일 또는 정상 메일로 분류하는 과정을 직접 수행해 보게 한다. 수업 활동 시 유의점 조건부확률의 기호나 용어는 도입하지 않고, 빈도를 활용하여 직관적으로 구해볼 수 있도록 한다. 평가의 주안점 확률을 구하여 스팸 메일 여부를 판단할 수 있는지 평가한다.
준비물	활동지

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	• 2~3명을 1개 모둠으로 구성 • 활동지 배부	
수행	모둠 활동1	• 스팸 메일의 유래, 뜻 알아보기 • 스팸 메일 필터의 전체 과정 이해하기 • 특정 단어 1개가 제목에 포함된 메일이 스팸 메일일 확률 구하기	포트폴리오 평가
	모둠 활동2	• 특정한 단어 1개가 제목에 포함된 메일이 스팸 메일일 확률 구하기 • 스팸 메일 필터의 원리를 다른 문제 상황에 적용하기	
정리		• 내용 정리	



스팸 메일 필터

년 월 일



스팸 메일이란?

1980년대 usenet이나 개인 이메일 사용자들이 지나친 광고글을 ‘스팸밍’이라고 부르기 시작하면서 스팸 메일이라는 단어가 생겼다. 스팸(spam) 메일의 반대인 정상적인 메일을 햄(ham)이라 부르기도 한다. 스팸 메일을 보내는 사람들은 제목에 엄청난 공을 들인다. 독창적이고 확인하고 싶도록 제목을 만들어야 수신자들이 스팸 메일을 열어보기 때문이다. 따라서 스팸 메일의 판정은 제목과 깊은 관련성을 갖는다. 스팸 메일 제목은 다양한 유형이 존재한다.



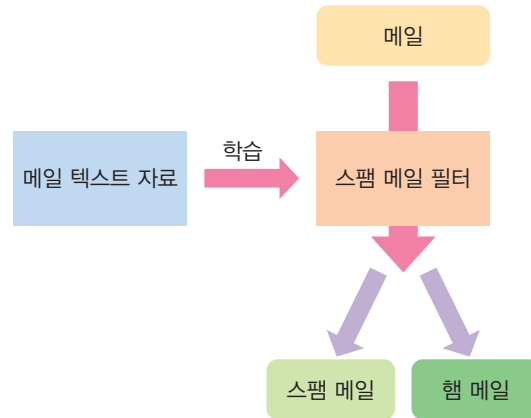
성적 어필형	▶불타는 금요일을 원하세요? ▶미성년자는 클릭 금지	궁금증 유발형	▶저기요.. ▶정말 아름다우십니다!
돈 관련형	▶만원으로 1억을 버는 방법 ▶이벤트 당첨을 축하합니다	관계 위장형	▶안녕하세요. 오랜만입니다. ▶RE: 안녕하세요. 특가 안내입니다.

인터넷 메일서비스에 접속해보면 스팸함에 자동으로 스팸 메일이 분류되어 있는 것을 확인할 수 있다.

받은편지함 21,142	한국철도공사 [한국철도]유연회원전환사원안내
다시 읽힐 항목	Sandeep Sharma Papers are invited for QUEST Journals - QUEST JOURNALS Inc. www.questjournals.org Dear Author, Publish your artic
중요편지함	하나투어 (광고) 인터컨티넨탈호텔 코엑스 최대 30% 할인 16만원대로 도심 속 특별한 힐링! - 제주여행 · 내륙여행 · 국내여행 · 다
보낸편지함	info@toughnut.info IOSR Refereed Journal - IOSR Journals (International Organization of Scientific Research) Dear Researcher "Wish You I
임시보관함 30	LightinTheBox.com narus4 고객님, 놀라운 깜짝 세일을 놓치지 마세요 - 남성 티셔츠&헝크 탑 - (CustomerName) 고객님, 놀라운 깜짝 세일을!
스팸함 26	IOSR Journal IOSR Best Impact Factor Journal - IOSR Journals www.submitpapersnow.com Dear Author, Sub: Publish paper in Best
카테고리	시공미디어 장기 미 이용 회원 개인정보 휴면전환 시행 안내 - 회원님의 계정이 휴면 상태로 전환 될 예정입니다. 안녕하세요. 회원님은
행아웃 미팅	Superbuy 11:11 Bonus: Order to Get ¥215 Coupon! ¥450 Off ¥500 & ¥100 Coupons Up For Grabs! First Come, First Served! 0 1
새 회의	Tata Capital Important Update on your Ex Gratia Payment - Tata Capital - Dear Sir/Madam, Greetings from Tata Capital. Tata Capita
회의 참여	

메일 서비스들이 자동으로 스팸 메일을 분류해서 스팸함으로 이동시키는 과정은 확률과 관계가 있다. 이번 시간에는 특정 단어가 포함된 조건의 이메일 중 스팸 메일일 확률을 계산하는 베이지안 스팸 필터에 대해 알아보자. 또, 자신의 스팸 메일함을 확인해 보고, 스팸 메일 제목이 갖는 주요 특징에는 어떤 것이 있는지 생각해 보자.

스팸 메일 필터는 스팸 메일과 햄 메일이 구분되어 있는 자료를 이용해 스팸 메일 구분법을 배울 수 있는 간단한 머신러닝 프로그램이라 할 수 있다.



다음은 지수가 하루 동안 받은 15개의 메일이다. 지수는 내용을 확인하고 각 메일이 스팸인지 아닌지 표시하여 스팸 메일 필터를 학습시키기 위한 학습용 자료를 만들었다.

연번	제목	스팸 여부
1	우리에게 당신의 비밀번호를 보내세요.	스팸
2	우리에게 확인 결과를 보내세요.	스팸
3	당신의 비밀번호를 확인하세요.	햄
4	해킹 여부를 확인해 주세요.	햄
5	우리에게 비밀번호를 보내세요.	스팸
6	우리에게 당신의 계정을 보내세요.	스팸
7	인공지능 수학 교과서 내용 확인 부탁드립니다.	햄
8	비밀번호가 초기화되었습니다.	햄
9	비밀번호 초기화 안내	스팸
10	비밀번호가 해킹되었습니다.	스팸
11	○○님 계정에서 해킹이 감지되었습니다.	스팸
12	데이터 백업을 위해 아이디와 비밀번호의 확인이 필요합니다.	스팸
13	난 당신 계정의 비밀번호를 알고 있다.	스팸
14	비밀번호의 보안단계가 낮아 해킹 위험도가 높습니다.	스팸
15	비밀번호 찾기 결과 확인	햄

1 지수가 스팸 메일과 햄 메일을 받을 확률을 각각 구해 보자.

$$\text{스팸 메일: } \frac{10}{15} = \frac{2}{3}, \text{ 햄 메일: } \frac{5}{15} = \frac{1}{3}$$

2 지수에게 제목에 “비밀번호”라는 단어가 포함된 메일이 올 확률을 구해 보자.

$$\frac{10}{15} = \frac{2}{3}$$

3 지수는 위의 데이터를 이용하여 “스팸 메일 필터”를 학습시켰다. 이 필터는 “비밀번호”라는 단어가 포함된 메일을 받았을 때, 이 메일을 어떻게 분류할까? 다음 표를 완성한 후 표를 이용하여 대답해 보자.

비밀번호가 제목에 포함된 총 메일의 수	10
비밀번호가 제목에 포함된 스팸 메일의 수	7
비밀번호가 제목에 포함된 햄 메일의 수	3

1) “비밀번호”라는 단어가 포함된 메일이 스팸 메일일 확률? $\frac{7}{10}$

2) “비밀번호”라는 단어가 포함된 메일이 햄 메일일 확률? $\frac{3}{10}$

3) 스팸 메일 필터의 최종 분류는?

스팸 메일일 확률은 $\frac{7}{10}$ 이고, 햄 메일일 확률은 $\frac{3}{10}$ 이므로 “비밀번호”라는 단어가 제목에 포함된 메일을 스팸 메일로 분류될 것이다.

4 스팸 메일 필터는 “확인”이라는 단어가 포함된 메일을 받았을 때, 이 메일을 어떻게 분류할까?

훈련용 데이터에서 “확인”이라는 단어가 제목에 포함된 메일은 총 6개이고, 그중에 2개는 스팸 메일, 4개는 햄 메일이다. 따라서 햄 메일일 확률이 $\frac{4}{6}$ 로 더 높기 때문에 햄 메일로 분류할 것이다.

5 지수는 대량의 메일 데이터를 이용하여 스팸 메일 필터를 학습시켰다. 이 데이터에서 2000개의 스팸 메일 중에 “대출”이라는 단어를 제목에 포함하는 것은 250개였다. 또 3000개의 햄 메일 중에 5개가 “대출”이라는 단어를 포함했다. 이 필터는 “대출”이라는 단어가 제목에 포함된 메일을 어떻게 분류하겠는가?

$\frac{250}{250+5}=0.9803\dots$ 이므로 스팸 메일로 분류할 것이다.

6 제목에 포함된 1개의 단어만으로 스팸 메일을 판정하는 것은 정확도가 떨어질 수 밖에 없다. 예를 들어 “대출”이라는 단어가 포함되었다고 모두 스팸 메일로 분류하면, 실제로 은행에서 온 대출 연장 안내 메일을 사용자가 받지 못하는 사고가 발생할 수 있다. 메일의 제목에는 여러 단어가 동시에 포함되어 있기 때문에 실제 스팸 메일 필터는 제목에 포함된 여러 단어를 함께 고려하여 스팸 메일 여부를 판단한다. 지수가 앞의 15개의 데이터로 학습시킨 스팸 메일 필터는 “비밀번호”와 “해킹” 두 단어가 모두 포함된 메일을 스팸 메일 필터가 어떻게 분류할지 다음 절차에 따라 알아보자.

1) 스팸 메일 필터 학습에 사용된 데이터를 보고 다음 확률을 구해 보자.

스팸 메일을 받을 확률	$\frac{2}{3}$
햄 메일을 받을 확률	$\frac{1}{3}$
스팸 메일 제목에 “비밀번호”가 포함될 확률	$\frac{7}{10}$
스팸 메일 제목에 “해킹”이 포함될 확률	$\frac{3}{10}$
햄 메일 제목에 “비밀번호”가 포함될 확률	$\frac{3}{5}$
햄 메일 제목에 “해킹”이 포함될 확률	$\frac{1}{5}$

2) 지수가 900개의 메일을 받는다고 했을 때, “비밀번호”와 “해킹”이 모두 제목에 포함된 스팸 메일은 몇 개 정도로 예상되는가? (단, A 단어와 B 단어가 출현할 확률은 독립적이라고 가정하자. 즉, A단어가 나타날 확률 $\times$ B단어가 나타날 확률 = A단어와 B단어가 동시에 나타날 확률)

$$900 \times \frac{2}{3} \times \frac{7}{10} \times \frac{3}{10} = 126$$

3) 지수가 900개의 메일을 받는다고 했을 때, “비밀번호”와 “해킹”이 모두 제목에 포함된 햄 메일은 몇 개 정도로 예상되는가? (단, A 단어와 B 단어가 출현할 확률은 독립적이라고 가정하자. 즉, A단어가 나타날 확률 $\times$ B단어가 나타날 확률 = A단어와 B단어가 동시에 나타날 확률)

$$900 \times \frac{1}{3} \times \frac{3}{5} \times \frac{1}{5} = 36$$

4) 2, 3을 이용하여 스팸 메일 필터가 “비밀번호”와 “해킹”이 모두 제목에 포함된 메일을 어떻게 분류할지 판단해 보자.

“비밀번호”와 “해킹”이 모두 포함된 메일 중에 스팸 메일의 비율을 생각하면 되므로

$$\frac{126}{126 + 36} = \frac{7}{9} = 0.77\cdots. \text{ 따라서 스팸 메일 필터는 스팸 메일로 분류할 것이다.}$$

7 스팸 메일 필터와 같은 방식으로 다음 표에 나와 있는 자료를 이용하여 예측을 하고자 한다.

연번	연령	태블릿 소지 여부	노트북 소지 여부
1	중학생	×	×
2	중학생	×	×
3	고등학생	×	○
4	대학생	×	○
5	대학생	○	○
6	대학생	○	×
7	고등학생	○	○
8	중학생	×	×
9	중학생	○	○
10	대학생	○	○
11	중학생	○	○
12	고등학생	×	○
13	고등학생	○	○
14	대학생	×	×

태블릿을 소지하고 있는 중학생은 노트북도 가지고 있을 것으로 예측하는가? 아니면 반대인가?

노트북을 소지할 확률	$\frac{9}{14}$
노트북을 소지하지 않을 확률	$\frac{5}{14}$
노트북을 소지한 사람이 중학생일 확률	$\frac{2}{9}$
노트북을 소지한 사람이 태블릿을 소지할 확률	$\frac{6}{9}$
노트북을 소지하지 않은 사람이 중학생일 확률	$\frac{3}{5}$
노트북을 소지하지 않은 사람이 태블릿을 소지할 확률	$\frac{1}{5}$

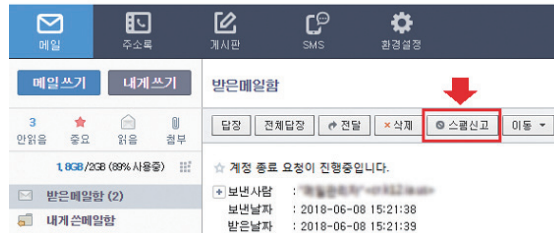
구하고자 하는 확률은

$$\frac{\frac{9}{14} \times \frac{6}{9} \times \frac{2}{9}}{\frac{9}{14} \times \frac{6}{9} \times \frac{2}{9} + \frac{5}{14} \times \frac{1}{5} \times \frac{3}{5}} = \frac{\frac{2}{21}}{\frac{2}{21} + \frac{3}{70}} = \frac{20}{29} = 0.689$$

이므로 노트북도 소지할 것으로 예측한다.

더 알아보기

“스팸 메일 중에 ‘비밀번호’라는 단어가 포함될 확률”과 같이 특정 조건 아래에서 사건이 일어날 확률을 조건부확률이라고 한다. 스팸 메일을 보내는 이들은 스팸 메일 필터를 피하기 위해 제목이나 메시지를 바꾸지만, 우리가 살펴본 조건부확률을 이용한 스팸 메일 필터 역시 지속적으로 개선될 수 있다. 여러분은 실제로 일반 메일함에 들어온 스팸 메일을 스팸이라고 표시하여 직접 스팸 메일함으로 옮겨본 기억이 있을 것이다. 이렇게 여러분이 직접 분류하는 결과가 스팸 메일 필터를 계속 학습시켜 정확도를 향상시킨다.



스팸 메일 필터에서 사용되는 분류 기법은 나이브 베이즈 분류라고 하는 머신러닝 중 지도학습의 일종이다. 나이브 베이즈는 간단하고 빠르며 정확도가 높은 모델이라는 장점이 있다. 그러나 정확도를 높이려면 빅데이터가 있어야 하며 특징 간에 독립성이 보장되어야 한다. 우리는 “비밀번호”와 “해킹”이라는 단어의 출현이 독립적이라고 가정하고 확률을 계산했지만, 실제로는 단어의 출현이 완전히 독립적이라고 볼 수 없다는 측면에서 한계가 존재한다. 예를 들면, “은행”이라는 단어가 포함되면 “계좌”라는 단어가 나타날 확률이 더 높을 것이기 때문이다. 최근에는 보다 정교한 머신러닝 기법을 이용하여 스팸 메일 필터의 정확도를 높이고 있지만, 기본적으로 나이브 베이즈 분류에 기초를 두고 있다.

지도의 주안점

- 조건부확률의 기호나 용어는 도입하지 않고, 빈도를 활용하여 직관적으로 구해 보는 활동을 통해 비형식적으로 이해할 수 있도록 지도한다.
- 작은 자료를 이용하여 직접 특정 단어 1개 또는 단어 2개가 제목에 포함된 메일을 스팸 메일 또는 정상 메일로 분류하는 과정을 직접 수행해 보게 한다.

핵심 발문

- 스팸 메일 중 “비밀번호”가 포함된 메일의 비율과 전체 메일 중 “비밀번호”가 포함된 메일의 비율은 같을까?
- 동전을 던졌을 때 앞면이 나올 확률은 $\frac{1}{2}$ 이다. 이때, 동전을 1000번 던지면 앞면이 몇 번 정도 나올 것으로 예상할 수 있을까?

교과서 연계 방안

- 조건부확률 관련하여 교과서를 활용할 수 있다.



관련 교과 역량

■ 추론

- 스팸 메일 필터의 학습과 예측 과정을 이해하는가?

■ 문제 해결

- 특정 단어 1개가 제목에 포함된 메일이 스팸 메일일 확률을 구할 수 있는가?
- 특정 단어 2개가 제목에 포함된 메일이 스팸 메일일 확률을 구할 수 있는가?
- 스팸 메일 필터의 원리를 다른 문제 상황에 적용할 수 있는가?

평가 결과 피드백

- 스팸 메일 필터와 관련된 문제 해결에 어려움을 겪는 경우 빈도를 중심으로 생각해볼 수 있게 발문한다.
- 스팸 메일 필터의 원리를 새로운 문제 상황에 적용하는 데 어려움을 겪는 학생의 경우, 스팸 메일 필터 상황과 새로운 문제 상황 사이의 대응 관계를 생각해 볼 수 있게 유도한다.

5. 모듈⑦

추세선을 이용한 예측은 어떻게 할까?

가 모듈 개관

성취기준	12인수03-04 자료의 경향성을 추세선으로 나타내고, 예측에 이용할 수 있다.
모듈 설계 의도	성취기준의 의미 자료를 산점도로 나타내어 상관관계를 판단하여 보고, 공학적 도구를 이용하여 자료의 경향성을 보여주는 일차함수 형태의 추세선을 찾아 결과 값을 예측해 볼 수 있다.
	학습 활동 계획 자료를 산점도로 나타내고 자료의 경향성을 직선으로 나타내 보게 한다. 상관관계가 있을 때 공학적 도구를 이용하여 자료의 경향성을 보여주는 일차함수 형태의 추세선을 찾고 새로운 x 의 값에 해당하는 y 의 값을 예측하는 활동을 한다.
	기대하는 학습 결과 자료의 경향성을 추세선을 이용하여 나타내고, 이를 이용하여 예측할 수 있다.
	모듈 자료에 대한 유의점 공학도구를 이용하여 구한 추세선의 식을 예측에 이용할 수 있음을 중점적으로 다룬다. 또한 자료의 경향성은 직선의 형태가 아닌 경우도 있으므로, 곡선의 경향성을 보이는 자료를 다루어 볼 수도 있다.
교과 역량	추론, 의사소통, 정보 처리
학습 요소	추세선
교수·학습 방법	매체 및 도구 활용 학습
교수·학습 방법 및 유의 사항	• 자료의 경향성은 관련된 두 자료를 산점도로 나타내고 측정값과 함숫값 사이의 오차를 최소화하는 추세선을 나타내는 두 변량 사이의 관계식 $y=ax+b$를 찾아 새로운 x의 값이 주어졌을 때, y의 값을 예측하는 과정을 이해하게 한다. • 추세선은 y의 측정값과 예측값 사이의 오차를 구하여 이를 최소화한다는 의미 정도로 간단히 다루고, 추세선을 구할 때 공학적 도구를 이용할 수 있다.
평가 방법	관찰 평가, 포트폴리오 평가
평가 방법 및 유의 사항	• 자료의 예측에 대한 평가는 확률 계산이나 관계식 등 수치 계산에 치중하지 않도록 하고, 자료의 경향성을 추세선으로 나타내는 활동과 확률을 예측에 이용하는 과정에 중점을 둔다.

차시	주제	교수·학습 활동	평가 활동
1	추세선을 이용한 예측	매체 및 도구 활용 학습 • 자료의 경향성을 잘 나타내 주는 추세선 선택하기 • 공학적 도구를 이용하여 추세선을 그리고 추세선의 식을 이용하여 예측하기 • SNS 이용자 수 자료에 대해 추세선을 이용하여 예측하고 실제값과 비교해 보기 • 모든 자료의 경향성을 직선 추세선으로 나타낼 수 있는 것이 아님을 이해하기	관찰 평가, 포트폴리오 평가 • 자료의 경향성을 직선으로 잘 나타내는가? • 공학적 도구를 활용하여 추세선을 구성할 수 있는가? • 추세선의 식을 이용하여 예측할 수 있는가? • 추세선을 이용한 예측 과정 및 결과에 대해 논리적으로 설명할 수 있는가? • 모든 자료의 경향성이 직선 형태가 아님을 이해하는가?

1차시 | 추세선을 이용한 예측

학습 목표	자료의 경향성을 추세선으로 나타내고 예측에 활용할 수 있다.
수업의 의도	수업 의도 공학적 도구를 이용하여 자료의 경향성을 보여 주는 추세선을 찾은 후, 새로운 x의 값에 해당하는 y의 값을 예측하는 경험을 할 수 있도록 설계하였다. 수업의 주안점 계산보다는 공학적 도구를 이용하여 구한 추세선의 식을 이용하여 예측에 활용하고 해석하는 부분을 중점적으로 다룬다. 또한 자료의 경향성은 직선의 형태가 아닌 경우도 있으므로, 곡선의 경향성을 보일 수도 있다는 점을 이해하게 한다. 수업 활동 시 유의점 컴퓨터나 스마트폰을 사용할 수 있는 수업 환경이 갖추어져야 한다. 평가의 주안점 수치 계산보다는, 자료의 경향성을 추세선으로 나타내고 예측에 이용하는 과정에 중점을 두고 평가를 한다.
준비물	활동지, 컴퓨터 또는 스마트폰

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	• 활동지 배분 및 수업 분위기 조성 • 2~3명을 1개 모둠으로 구성	
수행	모둠 활동	• 자료의 경향성을 잘 나타내 주는 추세선 선택하기 • 공학적 도구를 이용하여 추세선을 그리고 추세선의 식을 이용하여 예측하기 • SNS 이용자 수 자료에 대해 추세선을 이용하여 예측하고 실제값과 비교해 보기 • 모든 자료의 경향성을 직선 추세선으로 나타낼 수 있는 것이 아님을 이해하기	관찰 평가, 포트폴리오 평가
정리		• 학습 내용 정리	



추세선을 이용한 예측

년

월

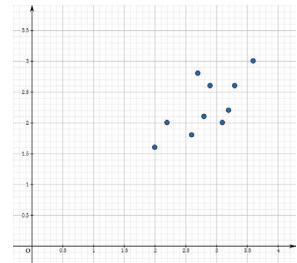
일



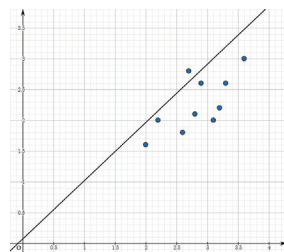
두 영화의 평점에 관한 자료이다. 다음 물음에 답해 보자.

영화 A와 영화 B는 유사한 성격의 영화이다. 오른쪽 표와 산점도는 영화 A와 영화 B에 대한 10명의 네티즌의 평점이다. 네티즌의 영화 A에 대한 평점과 영화 B에 대한 평점 사이에는 어떤 관계가 있는 것처럼 보인다. 한 네티즌이 영화 A에 평점 3.5를 주었을 때, 영화 B에 어떤 평점을 주었을지 예측하는 것이 가능할까?

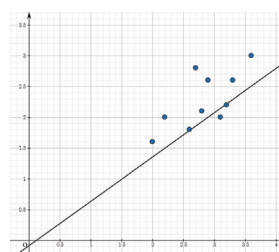
영화 A	영화 B
2	1.6
2.2	2
2.6	1.8
2.7	2.8
2.8	2.1
3.1	2.0
2.9	2.6
3.2	2.2
3.3	2.6
3.6	3



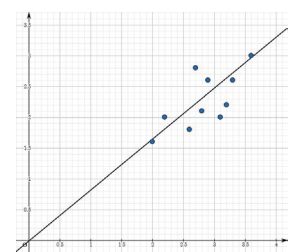
1 자료의 경향성을 좌표평면에서 일정한 직선 또는 곡선으로 나타내는 선을 추세선이라고 한다. 위에서 살펴본 두 영화의 평점에 관한 자료의 경향성을 하나의 직선 추세선으로 나타내려고 한다. 다음 세 가지 추세선 중 가장 적절하다고 생각하는 것이 무엇인지 고르고 이유를 이야기해 보자.



(a)



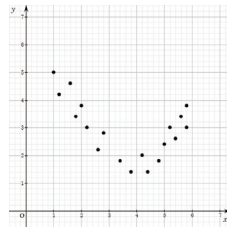
(b)



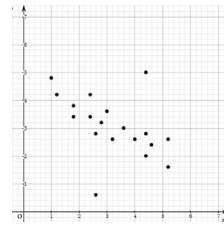
(c)

(a)는 너무 많은 자료가 아래에 위치하기 때문에 경향성을 잘 나타내는 직선이라고 보기 어렵다. 비슷한 이유로 (b)는 너무 많은 자료가 위에 위치하므로 부적절하다. 따라서 세 라인 중에 (c)가 가장 적절하다.

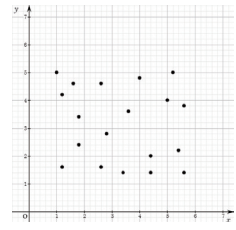
2 다음 세 산점도는 모두 그 경향성을 직선으로 잘 나타낼 수 있는가? 잘 나타낼 수 있는 것은 직선을 그려 나타내 보자.



(a)

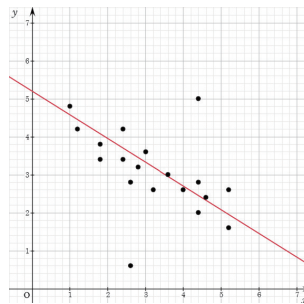


(b)



(c)

(a)는 그 경향성을 직선으로 나타내기 적합하지 않다. 오히려 (a)는 이차함수의 그래프와 같은 곡선이 그 경향성을 더 잘 나타낼 수 있을 것 같다. (b)의 경우는 위아래 양 끝에 위치한 점을 특별한 경우로 생각하면 기울기가 음인 직선 추세선으로 그 경향성을 다음과 같이 나타낼 수 있을 것 같다.

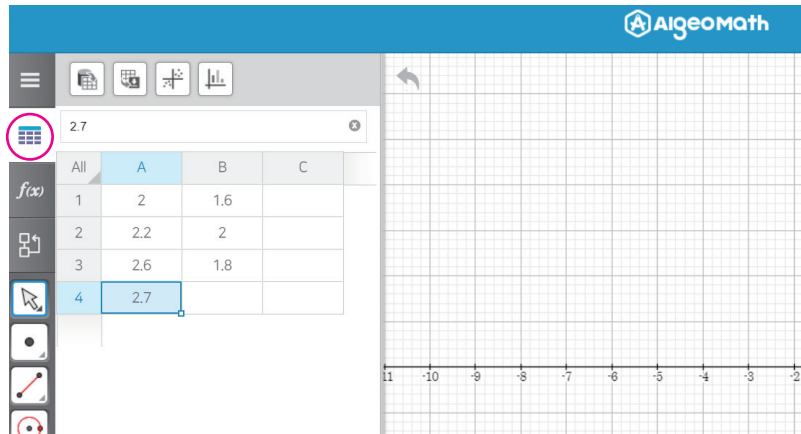


(c)의 경우도 기울기가 0인 추세선으로 경향성을 나타낼 수도 있으나 기울기가 0이기 때문에 관계성이 없다고 해석할 수 있다.

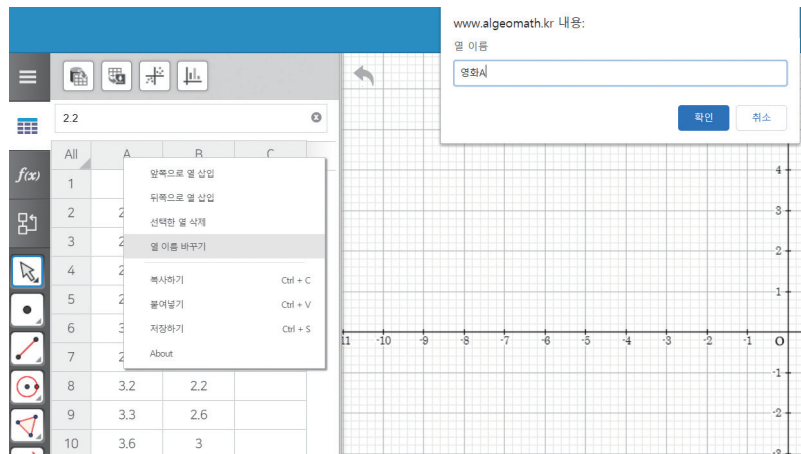
• 활동팁 학생들이 (c)를 직선으로 나타내는 것이 좋다고도 생각할 수도 있고 아니라고도 생각할 수도 있다. 정답이 없는 문제이기 때문에 답을 정해주기보다는 이에 대해 서로 토론해 보게 할 수 있다. 토론 이후 교사는 기울기가 0이고 y 절편이 y 값의 평균인 평행한 직선으로 나타내는 경우 직선으로 나타낼 수 있지만, 기울기가 없기 때문에 서로 관계성이 없음으로 해석할 수 있다고 정리할 수 있다.

 두 영화 평점 자료를 예시로 하여 알지오매스로 자료의 산점도를 그리고, 추세선을 그리는 방법을 알아보자.

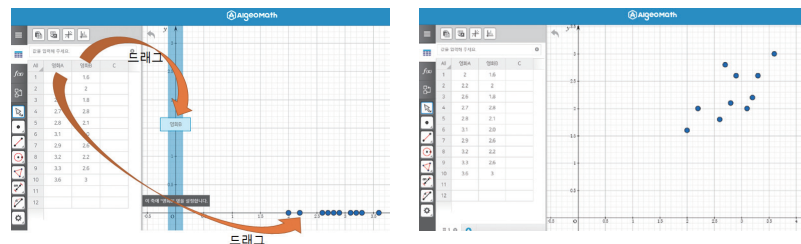
1) 왼쪽 메뉴에서 아이콘 를 클릭하면 표 도구가 나타난다. 표의 각 셀에 자료를 직접 입력할 수도 있으며, 엑셀이나 구글 스프레드시트 등에서 자료를 복사한 후 붙여넣기 할 수 있다.



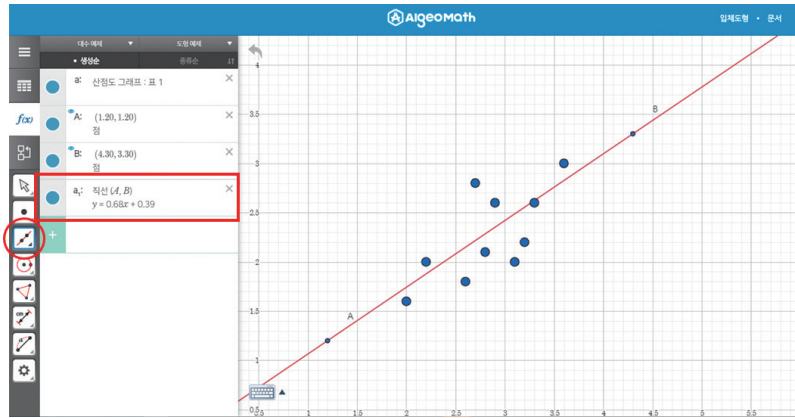
2) 각 열의 제목을 오른쪽 클릭한 후 “열 이름 바꾸기”를 선택하면 각 열의 이름을 설정할 수 있다.



3) 영화 A 열의 제목을 x 축에 드래그하고, 영화 B 열의 제목을 y 축에 드래그하면 산점도를 그릴 수 있다. 표 위의 아이콘 를 클릭해서 산점도를 그릴 수도 있다.



4) 작도 도구 중에 직선 도구 를 이용하면, 원하는 형태의 직선을 좌표평면에 그릴 수 있고, 대수창에서 직선의 방정식, 즉, 일차함수의 식을 확인할 수 있다.



1 위에서 두 영화 평점 사이의 관계를 추세선 $y=0.68x+0.39$ 로 나타냈다. 이 추세선에 의하면 영화 A에 평점 4점을 부여한 네티즌은 영화 B에 얼마의 평점을 부여할 것으로 예측되는가? 이러한 예측의 정확성은 어떻게 판단할 수 있을까?

$0.68 \times 4 + 0.39 = 3.11$ 이므로 3.11점을 줄 것으로 예측할 수 있다. 예측의 정확성을 판단하는 방법은 다양하게 있을 수 있는데, 이 경우 영화 A에 평점 4점을 준 사람들의 자료를 많이 구하여 그 사람들이 영화 B에 준 평점의 평균을 구한 다음 예측된 평점과 비교해 볼 수 있다.

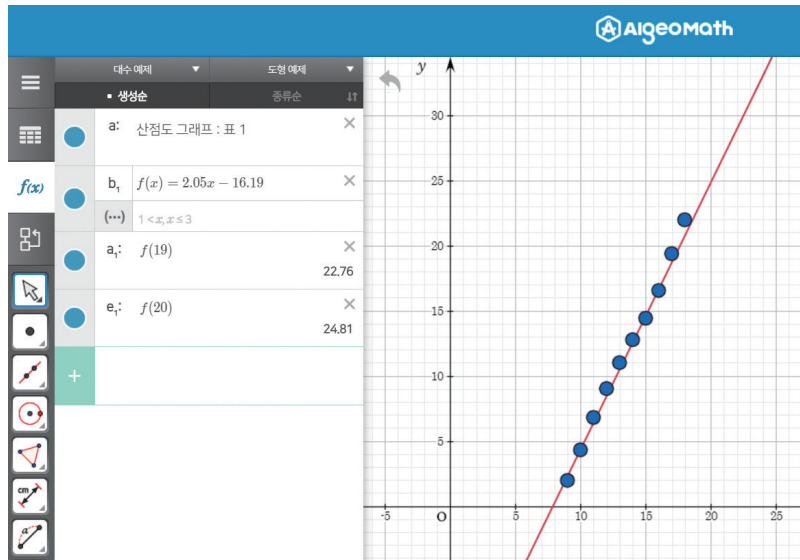


2 다음은 전 세계인이 사용하는 SNS 중 하나인 페이스북(facebook)의 10년(2009년 ~2018년)간 이용자 수를 정리한 것이다. 단, 09는 2009년, 18은 2018년을 의미한다.

연도	09	10	11	12	13	14	15	16	17	18
이용자 수 (단위: 억)	1.97	4.31	6.8	9.01	11.00	12.76	14.41	16.54	19.36	21.96

1) 알지오매스를 이용하여 자료에 알맞은 추세선의 식을 찾고, 페이스북의 2019년, 2020년 이용자 수를 예측해 보자. 또 친구의 추세선과 예측값을 비교해 보자.

	추세선의 방정식	2019년 이용자 수	2020년 이용자 수
나			
친구 ()			
친구 ()			



<http://me2.do/GUGrnyzS>

예를 들어, 추세선의 직선을 $f(x) = 2.05x - 16.19$ 로 설정할 수 있다. 이때, $f(19)$, $f(20)$ 의 값을 구해 보면 각각 22.76(억), 24.81(억)이다.

2) 인터넷 검색을 통해 실제 2019년과 2020년의 페이스북 이용자 수를 조사해 보고 자신의 예측과 비교해 보자.

실제 2019년 페이스북 이용자는 23.8억, 2020년 페이스북 이용자는 24.9억으로 예측이 어느 정도 맞았음을 알 수 있다.

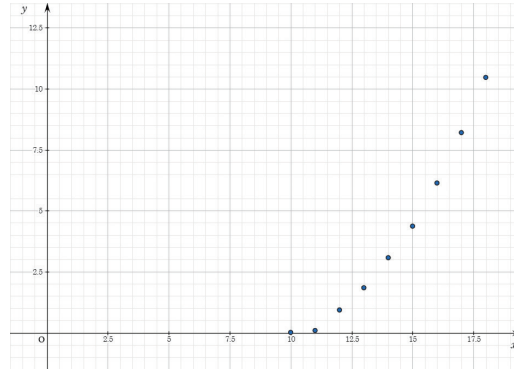


3 다음은 또 다른 SNS 중 하나인 인스타그램(instagram)의 9년(2010년~2018년)간 이용자 수를 정리한 것이다. 단, 10은 2010년, 18은 2018년을 의미한다.

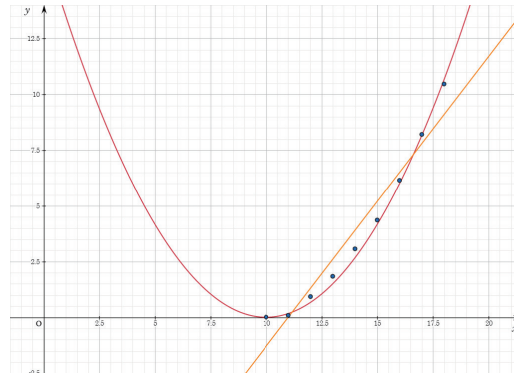
연도	10	11	12	13	14	15	16	17	18
이용자 수 (단위: 억)	0.01	0.09	0.93	1.84	3.07	4.36	6.15	8.21	10.47

1) 알지오매스를 이용하여 자료에 대한 산점도를 그려 보자.

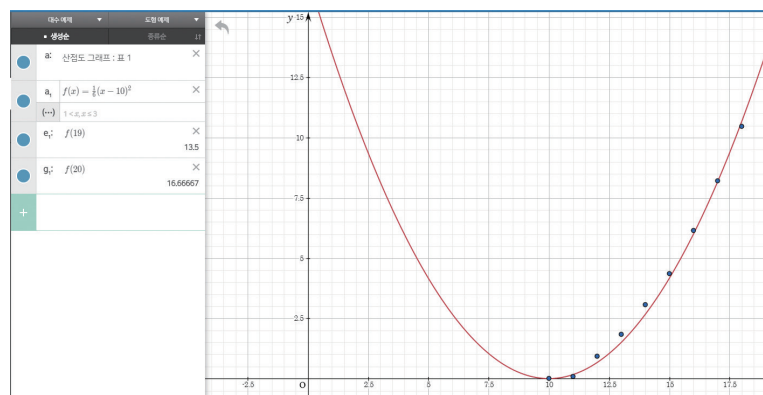
산점도를 그려 보면 다음과 같다.



2) 일차함수(직선)와 이차함수 중에 인스타그램 이용자 수의 경향성을 더 잘 나타낼 수 있는 것은 무엇인가? 일차함수 또는 이차함수 중 한 가지를 택해 알지오매스로 자료에 알맞은 추세선을 그리고, 인스타그램의 2019년, 2020년 이용자 수를 예측해 보자. 또 친구의 추세선과 예측값을 비교해 보자.



시간이 갈수록 증가량이 커지기 때문에 이차함수가 일차함수보다 더 적합한 것으로 보인다.



이차함수 $f(x) = \frac{1}{6}(x-10)^2$ 으로 경향성을 나타낼 수 있다. 이 함수를 이용하여 2019년과 2020년의 이용자 수를 예측하면 각각 13.5억명, 16.67억명이다.

3) 인터넷 검색을 통해 실제 2019년과 2020년의 인스타그램 이용자 수를 조사해 보고 자신의 예측과 비교해 보자.

인터넷 검색을 해 본 결과 인스타그램은 2020년에도 계속 10억명 정도의 이용자 수를 보유한 것으로 나타났다. 즉, 2018년에 10억명을 달성한 이후로는 이용자 수가 크게 늘지 않고 있는 것으로 보인다. 추세선을 이용한 예측이 항상 맞는 것은 아니라는 것을 알 수 있다.

지도의 주안점

- 예측이란 아직 일어나지 않은 상태나 현재 주어진 정보로 알 수 없는 상태를 추정해 보는 것이기 때문에, 정보를 더 구하거나 실제 사건이 일어난 후에 결과를 분석하여 예측과 비교해 보아 판단하는 것이라는 관점을 알려 준다.
- 계산보다는 공학적 도구를 이용하여 구한 추세선의 식을 이용하여 예측에 활용하고 해석하는 부분을 중점적으로 다룬다. 또한 자료의 경향성은 직선의 형태가 아닌 경우도 있으므로, 곡선의 경향성을 보일 수도 있다는 점을 이해하게 한다.

핵심 발문

- 모든 자료의 경향성을 직선 형태의 추세선으로 나타낼 수 있을까?
- 예측이 어려운 자료들은 어떤 특성을 가질까?



관련 교과 역량

- | | |
|---------|--|
| ■ 추론 | • 주어진 자료의 경향성을 가장 잘 설명할 수 있는 추세선을 선택할 수 있는가? |
| ■ 의사소통 | • 추세선을 이용하여 예측한 결과를 해석하고 설명할 수 있는가? |
| ■ 정보 처리 | • 공학적 도구를 이용하여 추세선을 긋고 추세선의 식을 찾아 활용할 수 있는가? |

평가 결과 피드백

- 추세선을 예측에 활용하지 못하거나 결과를 바르게 해석하지 못할 때는 추세선의 x , y 의 값이 자료와 어떻게 관계되는지 다시 한번 생각해 볼 수 있는 기회를 제공한다.
- 공학적 도구를 이용하여 추세선을 그리지 못하는 경우 직선 도구의 사용 절차에 대해 익힐 수 있도록 도움을 주고, 추세선의 식을 대수창에서 확인할 수 있음을 이해하게 한다.

IV

최적화

1. 영역의 개관

인공지능은 자료를 바탕으로 합리적인 의사 결정을 내리는 기술을 제공한다. 이때 주어진 자료에 가장 적합한 의사 결정 모델을 찾기 위해 함수를 만들고 최솟값 또는 최댓값을 찾아 문제를 해결할 수 있다.

가 2015 개정 수학과 교육과정 내용

1. 성취기준 및 학습 요소

최적화 영역에 제시된 성취기준 및 학습 요소는 다음과 같다.

Ⅰ 최적화와 의사 결정

[12인수04-01] 주어진 자료로부터 분류와 예측을 할 때, 오차를 표현할 수 있는 함수를 구성하는 원리와 방법을 이해한다.

[12인수04-02] 함수의 최솟값 또는 최댓값을 찾아 최적화된 의사 결정 방법을 이해한다.

[12인수04-03] 합리적 의사 결정과 관련된 인공지능 수학 탐구 주제를 선정하여 탐구를 수행한다.

관련 학습 요소

• 함수의 극한, 이차함수의 미분계수, 손실함수, 경사하강법

주요 학습 요소를 간단히 설명하면 다음과 같다.

손실함수

인공지능의 예측값과 실제값 사이의 차이를 손실 또는 오차라고 한다. 인공지능의 예측은 인공지능 모델에 포함된 가중치에 의해 결정되므로 손실은 가중치에 따라 달라진다고 볼 수 있다. 따라서 가중치에 따른 손실을 함수로 정의할 수 있는데, 이를 손실함수라 한다.

선형회귀에서 손실함수는 일반적으로 평균제곱오차를 이용하여 정의한다.

n 개의 자료와 일차함수 $y=ax$ 에 대한 손실함수 $E(a)$ 는

$$E(a) = \frac{1}{2n} \{ (y_1 - ax_1)^2 + (y_2 - ax_2)^2 + \cdots + (y_n - ax_n)^2 \}$$
으로 정의할 수 있다.

경사하강법

경사하강법은 인공지능에서 손실함수를 최적화시킬 때 사용되는 방법이다. 손실함수의 경사 반대 방향으로 일정한 보폭만큼 조금씩 움직여 가면서 손실함수가 최소가 되도록 하는 가중치를 찾아가는 방법이다. 교육과정에서 다루는 선형회귀에서 경사하강법의 절차는 다음과 같이 나타낼 수 있다.

- 1) 현재 x 의 값에 대한 미분계수 $f'(x)$ 를 구한다.
- 2) x 의 값을 $-\eta f'(x)$ 만큼 변화시킨다.(보폭: η)
- 3) 이와 같은 과정을 여러 번 반복한다.

2. 교수·학습 방법 및 유의 사항

최적화 영역의 교수·학습 방법 및 유의 사항은 다음과 같다.

- 함수의 극한 개념은 직관적 수준으로 다루고, 미분계수는 접선의 기울기로 도입한다.

〈인공지능 수학〉은 〈수학〉을 학습한 후 배울 수 있는 과목이므로 학생들이 미분을 학습하지 않을 수 있다. 함수의 극한, 평균변화율, 미분계수 등은 인공지능에서 손실함수를 최적화하는 데 핵심이 되는 개념이므로 일변수함수의 미분을 간단하게 도입한다.

함수의 극한과 미분계수의 정의는 경사하강법을 이해할 수 있는 정도로 직관적으로 도입한다.

- 오차를 표현할 수 있는 함수의 예로 손실함수를 간단히 다루며, 실제값과 예측값 사이의 오차로부터 손실함수를 유도한다.

분류에서의 손실함수는 활성화 함수 등의 개념을 필요로 하기 때문에, 상대적으로 간단한 선형회귀에서 실제값과 예측값 사이의 오차의 제곱의 평균을 이용하여 손실함수를 정의하는 과정을 중심으로 다룬다. 분류에서의 활성화 함수를 이용한 손실함수의 표현에 관한 내용은 읽기 자료 형태의 심화 학습으로 제시할 수 있다.

- 인공지능의 학습에는 지도학습, 비지도학습, 강화학습 등이 있음을 알고, 인공지능에서 학습 목표 중 하나가 손실함수를 최소화하는 것임을 이해하게 한다.

교육과정에서는 지도학습을 중점적으로 다루고 있지만, 인공지능의 종류에는 지도학습, 비지도학습, 강화학습이 있음을 이해하게 한다. 손실함수의 함수값, 즉 손실이 크다는 것은 인공지능의 정확도가 떨어진다는 것을 뜻하므로, 이를 최소화하는 것이 인공지능 학습의 목표임을 이해하게 한다.

- 최적화에서는 자료의 경향성을 나타내는 추세선을 예시로, 가장 적합한 모델을 찾기 위해 함수를 정의하는 과정과 이 함수의 최솟값 또는 최댓값을 찾는 과정을 이해하게 한다.

$y=ax+b$ 형태의 추세선을 직접 찾는 경우 손실함수가 $E(a, b)$ 형태의 이변수함수로 정의되는 어려움이 있다. 따라서 $y=ax$ 형태의 추세선을 찾는 상황에서 기울기 a 의 값에 따른 오차의 변화를 이차함수 $E(a)$ 로 정의하는 수준에서 간단히 다루고, 이 손실함수를 최소가 되도록 하는 a 의 값을 구하는 것이 최적화임을 이해하게 한다.

- 경사하강법은 이차함수 형태의 손실함수의 최솟값을 구하는 수준에서 간단히 다루고, 이때 공학적 도구를 이용할 수 있다.

$y=ax$ 형태의 추세선을 찾는 상황에서 손실함수는 이차함수 형태로 정의되어 학생들이 상대적으로 쉽게 이해할 수 있다. 함수의 극한과 미분계수의 정의를 직관적으로 소개하고, 이차함수 형태의 손실함수의 그래프에서 미분계수의 기하적 의미를 이용하여 함수의 값이 최소가 되는 방향으로 이동하는 방법을 이해하게 한다. 공학적 도구를 이용하여 경사하강법을 실습해 볼 수 있도록 한다.

- 경사하강법의 이동 단위인 학습률이 너무 크거나 작으면 학습이 제대로 이루어지지 않을 수 있음을 이해하게 한다.

경사하강법에서 학습률이 너무 작으면 학습이 지나치게 더디거나 학습이 완전히 이루어지지 않을 수 있다. 반면, 학습률이 너무 크면 최적화에 실패할 수 있다. 이러한 학습률의 특성을 학생들이 공학적 도구를 이용하여 직접 시뮬레이션해 보면서 이해할 수 있도록 한다.

- 합리적 의사 결정에서는 자율주행 자동차, 인공지능 가전, 재난 로봇, 스마트 팩토리 등 사물과 접목된 인공지능 기술 사례와 자동번역시스템, AI 비서, 바둑 프로그램, 추천시스템, 매칭 시스템, AI 보안 시스템, 챗봇 등 인공지능 기술의 사례 등을 다룰 수 있다.

인공지능을 활용하여 실생활의 편의성을 개선하거나 문제를 해결하는 사례를 통해 합리적 의사 결정과 관련된 수학 탐구 주제를 다양하게 경험하게 한다.

- 인공지능 수학 탐구에서는 인공지능을 이용하여 실생활 문제를 해결할 수 있는 다양한 아이디어를 탐색하는 탐구학습과 프로젝트 학습을 수행하게 한다. 이때, 창의적 아이디어와 연관된 수학이 무엇인지에 대해 발표하게 할 수 있다.

학생들이 직접 인공지능을 활용하여 실생활의 문제를 해결하는 아이디어를 탐구하고 발표하는 활동을 할 수 있다. 이때, 인공지능을 활용한 문제 해결 과정에서 수학적 개념이나 원리가 어떻게 사용되는지를 설명하게 함으로써 수학의 가치와 유용성을 이해할 수 있게 한다.

- 학생들의 수준에 맞추어 인공지능 기술을 직접 시연해 보거나 아이디어를 구현해 보게 하고, 이때 공학적 도구를 이용할 수 있다.

학생들이 인공지능 기술을 체험해 볼 수 있는 다양한 공학적 도구를 이용할 수 있다. 경사하강법과 관련해서는 프로그램의 코드 작성에 대한 부담을 갖지 않도록 하기 위하여 ‘자료’, ‘학습률’, ‘학습 횟수’, ‘학습을 위한 식’ 등 핵심적인 부분만을 수정하면서 ‘최적화’와 관련된 탐구와 실험을 할 수 있는 공학적 도구를 제공할 수 있다.

3. 평가 방법 및 유의 사항

최적화 영역의 평가 방법 및 유의 사항은 다음과 같다.

- 합리적 의사 결정 방법에 대한 이해나 아이디어를 평가할 때는 사례를 찾게 하거나 실제 자료를 이용하여 프로젝트를 수행하게 할 수 있으며, 이때 포트폴리오를 활용하여 과정 중심 평가를 할 수 있다.

자율주행 자동차, 인공지능 가전, 자동번역 시스템, 추천시스템 등과 같은 합리적 의사 결정에 관련된 인공지능 사례를 찾는 활동이나 학생들이 직접 인공지능을 활용하여 실생활의 문제를 해결하는 아이디어를 탐구하고 발표하는 프로젝트 활동을 할 수 있다. 이러한 프로젝트 활동에서의 산출물로 구성된 포트폴리오를 이용하여 과정 중심 평가를 실시할 수 있으며, 특히 인공지능을 활용한 문제 해결 과정에서 사용되는 수학적 개념이나 원리를 중심으로 평가한다.

2. 모듈⑧

예측의 정확도를 어떻게 높일까?

가 모듈 개관

성취기준	12인수04-01 주어진 자료로부터 분류와 예측을 할 때, 오차를 표현할 수 있는 함수를 구성하는 원리와 방법을 이해한다. 12인수04-02 함수의 최소값 또는 최대값을 찾아 최적화된 의사 결정 방법을 이해한다.
모듈 설계 의도	성취기준의 의미 주어진 텍스트, 이미지 자료로부터 분류와 예측을 할 때, 오차의 제곱의 합을 이용하여 손실함수를 구하는 원리와 방법을 이해한다. 이차함수 형태 손실함수의 최소값을 미분계수를 이용하여 찾을 수 있고, 최적화된 의사 결정 방법을 이해할 수 있다. 학습 활동 계획 추세선과 실제값 사이의 오차를 나타내고, 추세선의 기울기에 따라 변화하는 손실을 함수로 나타내는 활동으로 구성하였다. 미분계수를 공학적 도구를 이용하여 접선의 기울기로 직관적으로 도입하도록 구성하였다. 또한 그래프를 이용하여 경사하강법에 대해 스스로 고민해 보고, 공학적 도구를 이용하여 시뮬레이션하며 탐구할 수 있도록 구성하였다. 기대하는 학습 결과 인공지능의 학습이 손실함수의 최적화 과정임을 이해하고, 경사하강법의 원리를 이해한다. 모듈 자료에 대한 유의점 컴퓨터 또는 스마트폰을 사용할 수 있어야 한다. 미분은 직관적으로 도입하고, 계산보다는 최적화의 원리를 이해하는 데 수업의 목표를 둔다.
교과 역량	문제 해결, 추론, 정보 처리
학습 요소	함수의 극한, 이차함수의 미분계수, 손실함수, 경사하강법
교수·학습 방법	탐구 학습, 매체 및 도구 활용 학습
교수·학습 방법 및 유의 사항	- 함수의 극한 개념은 직관적 수준으로 다루고, 미분계수는 접선의 기울기로 도입한다. - 오차를 표현할 수 있는 함수의 예로 손실함수를 간단히 다루며, 실제값과 예측값 사이의 오차로부터 손실함수를 유도한다. - 인공지능의 학습에는 지도학습, 비지도학습, 강화학습 등이 있음을 알고, 인공지능에서 학습 목표 중 하나가 손실함수를 최소화하는 것임을 이해하게 한다.

	• 최적화에서는 자료의 경향성을 나타내는 추세선을 예시로, 가장 적합한 모델을 찾기 위해 함수를 정의하는 과정과 이 함수의 최솟값 또는 최댓값을 찾는 과정을 이해하게 한다. • 경사하강법은 이차함수 형태의 손실함수의 최솟값을 구하는 수준에서 간단히 다루고, 이때 공학적 도구를 이용할 수 있다. • 경사하강법의 이동 단위인 학습률이 너무 크거나 작으면 학습이 제대로 이루어지지 않을 수 있음을 이해하게 한다.
평가 방법	포트폴리오 평가
평가 방법 및 유의 사항	• 경사하강법에 대한 올바른 이해를 바탕으로 주어진 과제를 성실하게 해결하고 포트폴리오를 빠짐없이 기록하였는지를 중심으로 평가한다.

차시	주제	교수·학습 활동	평가 활동
1	예측의 정확도를 어떻게 평가할까?	탐구 학습 • 추세선의 오차 측정의 필요성 인식하기 • 추세선의 오차 표현 방법을 만들고 비교하기 • 손실함수 정의하기 • 손실함수의 최솟값 구하기 • 분류에서의 손실함수에 대해 더 알아보기	포트폴리오 평가 • 오차를 측정하는 합리적인 방법을 제안할 수 있는가? • 손실함수의 정의를 바르게 이해하는가? • 손실함수의 최솟값을 직접 구할 수 있는가?
2	미분과 기울기	매체 및 도구 활용 학습 • 구간 단속과 지점 단속 방식 비교하기 • 평균변화율에 대해 알아보기 • 순간변화율에 대해 알아보기 • 직관적인 방법으로 미분계수 구하기 • 공학적 도구를 이용하여 미분계수의 기하적 의미 이해하기 • 공학적 도구로 수치 미분 구하기	포트폴리오 평가 • 직관적인 방법을 통해 미분계수를 구할 수 있는가? • 공학적 도구를 이용하여 미분계수의 기하적 의미를 이해할 수 있는가? • 공학적 도구를 이용하여 수치 미분을 계산하는 과정을 이해하는가?
3~4	경사하강법	매체 및 도구 활용 학습 • 손실함수의 최솟값 지점으로 이동하기 위한 방향과 미분계수 사이의 관계 이해하기 • 경사하강법 알고리즘 찾기 • 직접 계산으로 경사하강법 적용해 보기 • 공학적 도구로 경사하강법 시뮬레이션하기 • 시뮬레이션을 통해 학습률에 대해 탐구하기	포트폴리오 평가 • 경사하강법에서 미분계수의 역할을 이해하였는가? • 경사하강법 알고리즘을 스스로 고안할 수 있는가? • 코드를 적절히 수정하여 새로운 상황에 적합한 경사하강법 시뮬레이션을 구성할 수 있는가? • 공학적 도구를 이용하여 경사하강법을 실험하고 학습률의 영향을 이해하였는가?

1차시 | 예측의 정확도를 어떻게 평가할까?

학습 목표	손실함수를 정의하여 예측의 정확도를 평가하는 방법을 이해한다.
수업의 의도	수업 의도 추세선의 오차를 측정해야 할 필요성을 인식한 후, 스스로 오차를 측정하는 다양한 방법을 생각하고 비교해 볼 수 있도록 구성하였다. 정의를 이용하여 손실함수를 직접 구하고, 최솟값을 구해 볼 수 있도록 설계하였다. 수업의 주안점 오차를 표현하는 다양한 방법이 있으나, 계산 등의 편의를 위하여 평균제곱오차를 주로 이용함을 이해하게 한다. 실제 자료에 대해 손실함수를 구해 보고 직접 최솟값을 찾아볼 수 있게 한다. 수업 활동 시 유의점 분류에서의 손실함수는 심화 학습 자료로 제공하여 직관적인 수준에서 이해할 수 있게 한다. 평가의 주안점 손실함수의 필요성을 인식할 수 있는지, 정의를 이용하여 손실함수를 정확히 구할 수 있는지를 중점적으로 평가한다.
준비물	활동지

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	• 2~3명을 1개 모둠으로 구성 • 활동지 배분 및 수업 분위기 조성	
수행	모둠 활동	• 추세선의 오차 측정의 필요성 인식하기 • 추세선의 오차 표현 방법을 만들고 비교하기 • 손실함수 정의하기 • 손실함수의 최솟값 구하기 • 분류에서의 손실함수에 대해 더 알아보기	포트폴리오 평가
정리		• 학습 내용 정리	



예측의 정확도를 어떻게 평가할까?

년

월

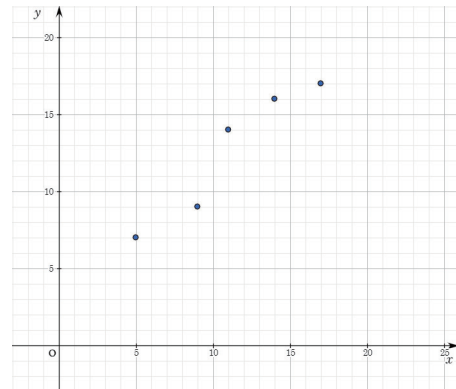
일



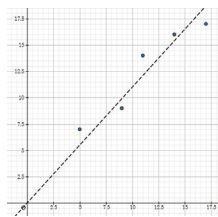
자료를 가장 잘 나타내 주는 추세선을 어떻게 찾을 수 있을까? 다음 물음에 답해 보자.

일반적으로 날씨가 더우면 아이스크림의 판매량은 증가할 것으로 예상할 수 있다. 실제로 다음은 어떤 슈퍼마켓의 온도에 따른 아이스크림 판매량을 조사한 것이다.

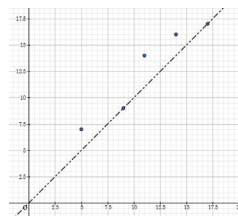
온도(도)	판매량(개)
5	7
9	9
11	14
14	16
17	17



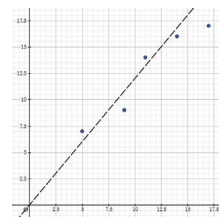
1 철수, 민수, 영희는 원점을 지나는 일차함수 $y=ax$ 꼴로 추세선을 다음과 같이 그렸다.



$$y = \frac{11}{10}x$$



$$y = x$$



$$y = \frac{6}{5}x$$

추세선이 자료의 경향성을 정확히 나타낼 수 있어야 추세선을 이용한 예측이 정확하다고 말할 수 있을 것이다. 위 그림에서 $y = \frac{11}{10}x$, $y = x$, $y = \frac{6}{5}x$ 중 자료의 경향성을 가장 잘 나타내는 $y = ax$ 꼴 추세선은 무엇이라고 생각되는가?

$y = \frac{11}{10}x$ 또는 $y = \frac{6}{5}x$ 일 것 같은데 정확히 알 수 없다. 기준을 세워서 수학적으로 찾을 필요가 있다.

2 자료가 완전한 일차함수를 따르는 것이 아니므로 모든 점을 지나는 직선은 존재하지 않는다.
따라서 추세선과 실제값 사이의 오차를 가장 작게 만드는 추세선을 찾아야 한다.

1) 추세선과 실제 자료 사이의 오차를 나타낼 수 있는 방법을 생각해 보자.

2) 1)에서 생각한 방법을 이용하여 추세선 $y = \frac{11}{10}x$ 와 $y = x$ 각각의 오차를 계산하고 비교해 보자.

학생들이 여러 가지 방안을 제시할 수 있다. 다양한 방안을 서로 비교하게 함으로써 장단점을 파악하게 한다.

(i) 거리 이용

추세선과 자료 사이의 거리를 이용하는 것이다.

$y = x$ 의 경우

온도	판매량	예측값	오차
5	7	5	$ 7 - 5 = 2$
9	9	9	$ 9 - 9 = 0$
11	14	11	$ 14 - 11 = 3$
14	16	14	$ 16 - 14 = 2$
17	17	17	$ 17 - 17 = 0$

이므로 오차의 총합은 7이다.

$y = \frac{11}{10}x$ 의 경우

온도	판매량	예측값	오차
5	7	$\frac{55}{10}$	$ 7 - 5.5 = 1.5$
9	9	$\frac{99}{10}$	$ 9 - 9.9 = 0.9$
11	14	$\frac{121}{10}$	$ 14 - 12.1 = 1.9$
14	16	$\frac{154}{10}$	$ 16 - 15.4 = 0.6$
17	17	$\frac{187}{10}$	$ 17 - 18.7 = 1.7$

이므로 오차의 총합은 6.6이다.

따라서 추세선 $y = \frac{11}{10}x$ 가 $y = x$ 보다는 오차가 작다.

(ii) 차의 제곱 이용

추세선과 자료 사이의 차를 제공하는 것이다.

$y=x$ 의 경우

온도	판매량	예측값	오차
5	7	5	$(7-5)^2=4$
9	9	9	$(9-9)^2=0$
11	14	11	$(14-11)^2=9$
14	16	14	$(16-14)^2=4$
17	17	17	$(17-17)^2=0$

이므로 오차의 총합은 17이다.

$y=\frac{11}{10}x$ 의 경우

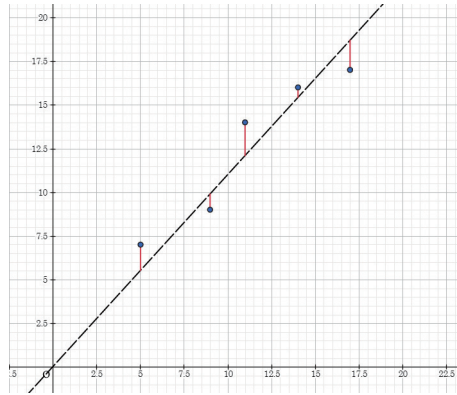
온도	판매량	예측값	오차
5	7	$\frac{55}{10}$	$(7-5.5)^2=2.25$
9	9	$\frac{99}{10}$	$(9-9.9)^2=0.81$
11	14	$\frac{121}{10}$	$(14-12.1)^2=3.61$
14	16	$\frac{154}{10}$	$(16-15.4)^2=0.36$
17	17	$\frac{187}{10}$	$(17-18.7)^2=2.89$

이므로 오차의 총합은 9.92이다.

따라서 추세선 $y=\frac{11}{10}x$ 가 $y=x$ 보다는 오차가 작다.



추세선을 이용한 예측값과 실제값 사이의 오차를 나타낸 그래프이다. 다음을 살펴보자.



결국 빨간 선분들의 길이를 이용하면 추세선과 실제값 사이의 오차를 표현할 수 있을 것이다. 추세선을 일차함수 $y=ax$, 실제 자료의 좌표를 차례로 $(x_i, y_i) (i=1, 2, 3, 4, 5)$ 라 하면, 오차의 총합은 $\sum_{i=1}^5 |y_i - ax_i|$ 로 나타낼 수 있다. 그러나 이렇게 오차를 나타내면 자료의 개수가 늘어날수록 오차도 커진다. 따라서 자료의 수가 서로 다른 경우 오차의 정도를 비교하기 어렵다. 따라서 평균적인 오차를 나타내기 위해 자료의 수로 나누어 주면 $\frac{1}{5} \sum_{i=1}^5 |y_i - ax_i|$ 로 나타낼 수 있다. 일반적으로 미분⁸⁾을 쉽게 하도록 하기 위하여 절댓값 대신 오차를 제공하고 $\frac{1}{2}$ 을 곱한 식인 $\frac{1}{2} \cdot \frac{1}{5} \sum_{i=1}^5 (y_i - ax_i)^2$ 으로 평균적인 오차를 표현한다. $\frac{1}{2} \cdot \frac{1}{5} \sum_{i=1}^5 (y_i - ax_i)^2$ 에서 x_i, y_i 는 상수이므로 이는 추세선의 기울기 a 에 따라 값이 변화하는 함수다. 따라서 이를 예측값과 실제값 사이의 오차를 나타내는 손실함수 또는 오차함수라 부른다.

❖ 추세선을 이용한 예측에서 손실함수

n 개의 자료와 일차함수 $y=ax$ 에 대한 손실함수 $E(a)$ 는

$$E(a) = \frac{1}{2n} \{ (y_1 - ax_1)^2 + (y_2 - ax_2)^2 + \cdots + (y_n - ax_n)^2 \}$$

7) $\sum_{i=1}^5 f(i) = f(1) + f(2) + f(3) + f(4) + f(5)$ 이다.

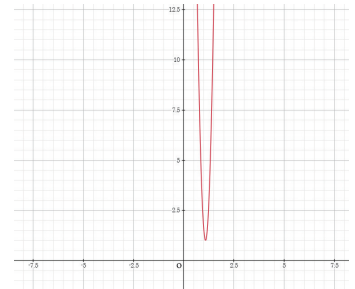
8) 미분에 대해서는 다음 차시에서 배운다.

1 일차함수 $y=ax$ 꼴 추세선을 예측 모델로 활용할 때, 아이스크림 자료에 대한 손실함수 $E(a)$ 를 구해 보자.

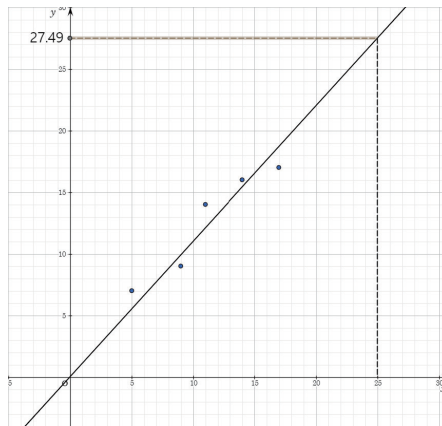
$$\begin{aligned} E(a) &= \frac{1}{10} \sum_{i=1}^5 (y_i - ax_i)^2 \\ &= \frac{1}{10} \{ (5a-7)^2 + (9a-9)^2 + (11a-14)^2 + (14a-16)^2 + (17a-17)^2 \} \\ &= \frac{1}{10} (712a^2 - 1566a + 871) \end{aligned}$$

2 1에서 구한 손실함수가 최소가 되도록 하는 a 의 값을 구하고, 그것을 이용하여 온도가 25도인 날의 아이스크림의 예상 판매량을 예측해 보자.

손실함수는 a 에 대한 이차함수로 표현되며 그 그래프는 오른쪽 그림과 같다. 완전제곱을 이용하여 계산하면 $a = \frac{783}{712} = 1.0997\cdots$ 일 때 $E(a)$ 가 최소가 된다는 것을 알 수 있다.



이제, 우리가 찾은 모델을 원자료와 함께 그래프에 나타내면 다음과 같다.



우리가 찾은 모델을 이용하면 온도가 25도인 날의 아이스크림의 예상 판매량은 27.49개라 할 수 있다.

분류에서의 손실도 유사한 방식으로 수치화할 수 있다.

예를 들어, MNIST 손글씨 숫자 인식 예제의 '2'를 쓴 손글씨 이미지를 2개의 서로 다른 신경망 A, B가 예측한 값을 각각 예측값 A, 예측값 B라 하자.

	'0'	'1'	'2'	'3'	'4'	'5'	'6'	'7'	'8'	'9'
실제값	0	0	1	0	0	0	0	0	0	0
예측값 A	0.1	0.05	0.6	0	0.05	0.1	0	0.1	0	0
예측값 B	0	0.01	0.6	0.16	0.02	0.2	0.01	0	0	0

정답은 '2'이므로 실제값은 '2'에만 1이 코딩되어 있다.

예측은 확률로 해석할 수 있다. 예를 들어 신경망 A는 손글씨가 숫자 '0'일 확률은 0.1, '1'일 확률은 0.05, '2'일 확률은 0.6, ...으로 계산한 것이고, '2'일 확률이 가장 높게 활성화되었으므로 '2'로 예측한 것이다. 신경망 B도 역시 동일하게 '2'일 확률이 0.6으로 가장 높으므로 '2'로 예측한 것이다. 이때, 추세선에서 손실함수를 정의할 때와 유사하게 평균제곱오차를 이용하면 두 신경망의 예측의 정확도를 수치화하여 비교할 수 있다.

예를 들어, 예측값 A와 실제값의 각 확률 사이의 오차를 제곱하여 평균을 구하면

$$\frac{1}{2} \cdot \frac{1}{10} \{ (0.1-0)^2 + (0.05-0)^2 + (0.6-1)^2 + (0-0)^2 + (0.05-0)^2 + (0.1-0)^2 + (0-0)^2 + (0.1-0)^2 + (0-0)^2 + (0-0)^2 \} = 0.00975$$

예측값 B의 경우는

$$\frac{1}{2} \cdot \frac{1}{10} \{ (0-0)^2 + (0.01-0)^2 + (0.6-1)^2 + (0.16-0)^2 + (0.02-0)^2 + (0.2-0)^2 + (0.01-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \} = 0.01131$$

따라서 평균제곱오차를 이용하면 예측값 A의 오차가 예측값 B의 오차보다 더 작기 때문에 동일하게 '2'를 예측했어도 신경망 A의 예측이 상대적으로 정확하다고 판단할 수 있다.

시도의 주안점

- 오차를 표현하는 다양한 방법이 있으나, 계산 등의 편의를 위하여 평균제곱오차를 주로 이용함을 이해하게 한다. 실제 자료에 대해 손실함수를 구해 보고 직접 최솟값을 찾아볼 수 있게 한다.
- 최적화한 추세선을 이용하여 예측해 보는 경험을 하게 한다.

핵심 발문

- 오차를 숫자로 표현하려면 어떻게 해야 할까?
- 이차함수는 어디에서 최솟값을 가질까?



관련 교과 역량

- | | |
|---------|--------------------------------|
| ■ 추론 | • 오차를 측정하는 합리적인 방법을 제안할 수 있는가? |
| ■ 문제 해결 | • 손실함수의 최솟값을 직접 구할 수 있는가? |

평가 결과 피드백

- 오차를 측정하는 방법을 생각해 내지 못할 때는, 추세선이 자료의 경향성을 얼마나 잘 나타내는지를 어떤 수치로 비교할 수 있을지 생각해 보게 한다.
- 손실함수의 최솟값을 구하지 못하는 경우, 이차함수의 성질에 관한 내용을 상기시킨다.

2차시 | 미분과 기울기

학습 목표	미분계수의 의미를 이해한다.
수업의 의도	수업 의도 경사하강법을 이해하기 위한 선수 학습 요소로 미분계수의 기하적 의미를 직관적으로 이해할 수 있도록 설계하였다. 수업의 주안점 매우 작은 구간에서의 평균변화율이 순간변화율임을 이해하게 한다. 직관적인 극한을 이용하여 이차함수의 미분계수를 구할 수 있도록 하였다. 공학적 도구를 이용하여 미분계수가 접선의 기울기라는 것을 이해할 수 있도록 설계하였다. 수업 활동 시 유의점 극한은 충분히 작은 수를 이용하여 직관적으로 다루고, 미분계수의 기하적 의미는 공학적 도구를 이용하여 이해할 수 있게 한다. 평가의 주안점 미분계수의 계산보다는 미분계수의 기하적 의미를 정확히 이해했는지를 중점적으로 평가한다.
준비물	활동지, 컴퓨터, 태블릿 PC 또는 스마트폰

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	• 2~3명으로 모둠 구성 • 활동지 배분 및 수업 분위기 조성	
수행	모둠 활동	• 구간 단속과 지점 단속 방식 비교하기 • 평균변화율에 대해 알아보기 • 순간변화율에 대해 알아보기 • 직관적인 방법으로 미분계수 구하기	포트폴리오 평가
		• 공학적 도구를 이용하여 미분계수의 기하적 의미 이해하기 • 공학적 도구로 수치 미분 구하기	
정리		• 학습 내용 정리	



미분과 기울기

년

월

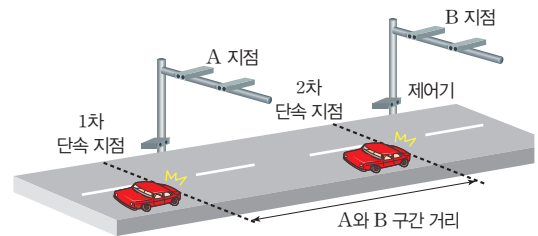
일



미분은 그래프에 대한 중요한 정보를 제공하는 것으로, 인공지능에서 손실함수의 최솟값을 찾아 예측의 정확도를 높이는 데 유용하게 활용되는 수학적 개념이다. 다음 물음에 답해 보자.

자동차 과속 단속 방식에는 일정 구간에서의 평균 속도를 측정하는 구간 단속과 특정 지점에서의 순간적인 속도를 측정하는 지점 단속이 있다. 구간 단속은 특정 구간의 시작 지점과 종료 지점에 단속 카메라를 설치하고 차량이 구간의 시작 지점과 종료 지점을 통과하는 시간을 측정하여 차량이 두 지점을 통과할 때의 평균 속도를 측정하여 규정 속도와 비교한다. 지점 단속은 단속 카메라가 설치되어 있는 지점을 통과할 때의 짧은 순간의 속도를 측정하여 규정 속도 위반 여부를 확인하는 방식이다.

$$\frac{\text{A와 B 구간 거리}}{\text{B 지점 단속 시간} - \text{A 지점 단속 시간}} = \text{구간 평균 속도}$$



1 구간 단속 방식과 지점 단속 방식의 장단점에 대해 이야기해 보자.

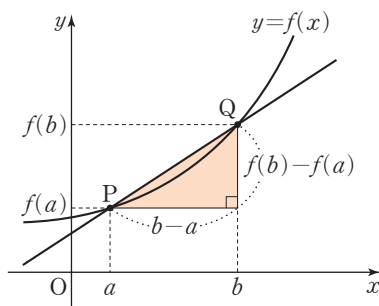
- 구간 단속: 비교적 긴 구간에서 과속을 방지하는 효과가 있다. 과속을 했더라도 종료 지점의 단속 카메라를 통과하기 전에 갓길 등에서 대기하고 있다 통과하면 단속을 피할 수 있다는 헛점이 있다.
- 지점 단속: 교통사고가 많이 발생하는 지점이나 어린이 보호 구역 등에 설치하여 해당 지점 근처에서 차량의 속도를 줄이도록 함으로써 사고를 예방하는 효과를 얻을 수 있다. 그러나 해당 지점을 통과하고 나면 운전자들이 다시 쉽게 과속할 수 있다는 점이 단점이다.



평균변화율에 대해 알아보자.

함수 $y=f(x)$ 의 그래프 위의 두 점 $P(a, f(a))$, $Q(b, f(b))$ 에 대하여 $\frac{f(b)-f(a)}{b-a}$ 를 x 의 값이 a 에서 b 까지 변할 때 함수 $y=f(x)$ 의 평균변화율이라고 한다. 평균변화율은 구간 단속에 해당하는 개념으로 볼 수 있다.

1 함수의 그래프에서 평균변화율이 갖는 기하적 의미는 무엇인지 생각해 보자.



두 점 $P(a, f(a))$, $Q(b, f(b))$ 를 지나는 직선의 기울기이다.

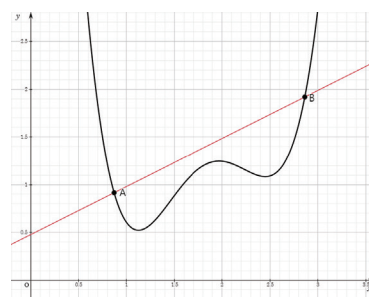
2 함수 $f(x) = -x^2 + 2x + 1$ 에서 x 의 값이 1에서 3까지 변할 때의 평균변화율을 구해 보자.

$$\frac{f(3)-f(1)}{3-1} = \frac{-4}{2} = -2$$



순간변화율에 대해 알아보자.

평균변화율은 함수가 전체적으로 증가 또는 감소하고 있음을 나타내 줄 수 있지만 세밀한 부분까지 알려 줄 수 없다. 예를 들어 오른쪽 함수는 A에서 B까지 평균적으로는 증가했다고 볼 수 있지만, 자세히 살펴보면 증가와 감소하는 구간이 모두 포함되어 있다. 따라서 함수의 세밀한 변화를 확인하기 위해서는 아주 작은 구간에서 함수의 증감을 확인할 필요가 있는데, 이를 순간변화율 또는 미분계수라 하며, 지점 단속에 대응되는 개념이라 할 수 있다.



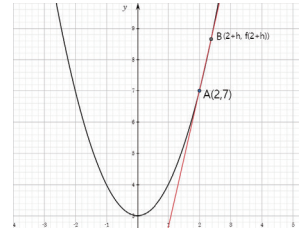
1 다음은 이차함수 $f(x) = x^2 + 3$ 의 $x=2$ 에서의 미분계수를 구하는 과정이다.

2에서 $2+h$ 까지의 평균변화율은

$$\frac{f(2+h)-f(2)}{(2+h)-2} = \frac{(2+h)^2-2^2}{h} = \frac{h^2+4h}{h} = h+4$$

이다. h 가 0에 가까운 아주 작은 수라고 생각하면

미분계수는 4가 될 것이다.



위 과정을 참고하여 이차함수 $f(x) = ax^2 + bx + c$ 의 $x=k$ 에서의 미분계수는 $2ak + b$ 임을 보여 보자.

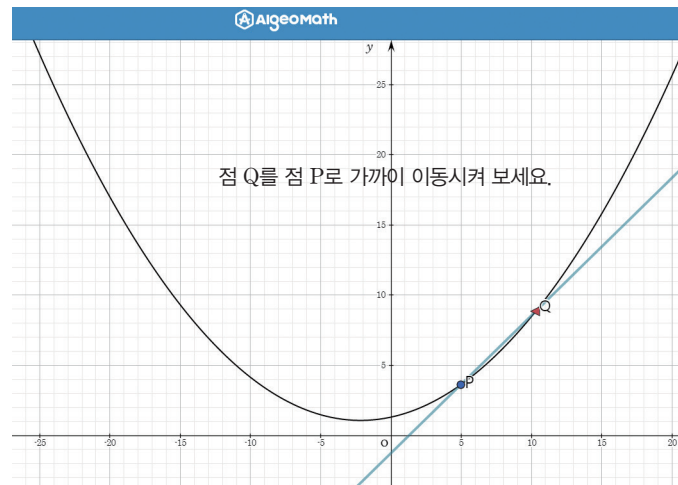
k 에서 $k+h$ 까지의 평균변화율은

$$\frac{f(k+h)-f(k)}{(k+h)-k} = \frac{a(k+h)^2+b(k+h)-ak^2-bk}{h} = \frac{ah^2+2akh+bh}{h} = ah+2ak+b$$

이다. h 가 0에 가까운 아주 작은 수라고 생각하면 미분계수는 $2ak + b$ 이다.



2 AI 웹실험실의 '미분계수의 기하적 의미'에서 다음과 같은 그래프를 확인할 수 있다.



점 Q를 점 P로 가깝게 이동시키면서 직선 PQ의 기울기를 살펴보자. 점 P에서의 미분계수가 갖는 기하적 의미는 무엇이라고 생각하는가?

함수의 그래프를 점 P에서 스치듯이 지나가는 직선의 기울기, 즉, 접선의 기울기이다.

수치 미분에 대해 알아보자.

미분계수는 원래 두 점 $(a, f(a))$, $(a+h, f(a+h))$ 의 x 좌표 사이의 간격 h 가 한없이 0으로 가까이 갈 때의 값이다. 컴퓨터는 이와 같이 ‘한없이 가까이 가는 수’를 표현할 수 없다. 따라서 머신러닝에서 미분계수를 구할 때는 수치 미분이라는 방법을 사용한다. 수치 미분이란 간격 h 를 0에 한없이 가까이 가는 수가 아니라 0.001과 같이 아주 작은 수로 정해서 미분계수의 근삿값을 수치적으로 구하는 것을 말한다.

활동

1 왼쪽은 알지오매스에서 $f(x) = x^2 + 2x$ 의 $x=2$ 에서의 미분계수를 수치 미분을 이용하여 구하는 코드와 실행 결과이고, 오른쪽은 알지오매스의 대수창을 이용하여 미분계수를 구한 것이다. 오차가 매우 작은 것을 확인할 수 있다.



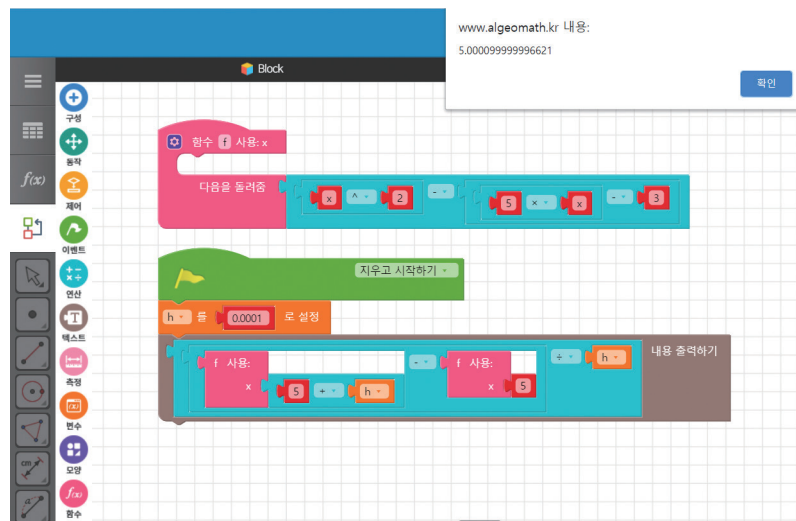
수치 미분을 이용한 근삿값

	대수 예제	도형 예제
	생성순	종류순
	a ₁	$f(x) = (x)^2 + 2x$
	(...)	$1 < x, x \leq 3$
$f(x)$	d ₁	$f'(2)$
		6

실제 미분계수

(AI 웹실�험실)의 ‘수치 미분’에서 주어진 코드를 수정하여 함수 $f(x) = x^2 - 5x + 3$ 의 $x=5$ 에서의 미분계수를 수치 미분을 이용하여 구해 보자.

다음과 같은 결과를 얻을 수 있다.



시도의 주안점

- 매우 작은 구간에서의 평균변화율이 순간변화율임을 이해하게 한다.
- 직관적인 극한을 이용하여 이차함수의 미분계수를 구할 수 있도록 한다.
- 공학적 도구를 이용하여 미분계수가 접선의 기울기라는 것을 이해할 수 있도록 한다.

핵심 발문

- 점 Q가 점 P에 충분히 가까이 갔을 때 직선 PQ의 기울기와 점 P에서의 미분계수는 어떤 관계가 있을까?

교과서 연계 방안

- 미분계수에 관한 심화 학습을 할 때 수학 I 교과서를 활용할 수 있다.



관련 교과 역량

- 추론
 - 미분계수의 기하적 의미를 바르게 추론하는가?
- 정보 처리
 - 공학적 도구를 이용하여 수치 미분을 계산할 수 있는가?

평가 결과 피드백

- 미분계수를 구하는 데 어려움을 겪는 경우 평균변화율을 먼저 구한 뒤 순간변화율을 구하도록 유도한다.
- 공학적 도구를 이용한 활동에 어려움을 겪는 경우 구체적인 사례를 이용하여 도구의 사용법을 다시 한번 안내한다.

3~4차시 | 경사하강법

학습 목표	경사하강법을 이용하여 손실함수의 최솟값을 구하는 과정을 이해한다.
수업의 의도	수업 의도 경사하강법의 원리를 이해하고, 공학적 도구를 이용한 시뮬레이션을 통해 경사하강법에 대해 탐구할 수 있도록 설계하였다. 수업의 주안점 미분계수의 부호와 손실함수의 최솟값 지점으로 이동하기 위한 방향 사이의 관계를 스스로 이해할 수 있게 한다. 2, 3회 정도 경사하강법을 직접 계산해 보도록 한다. 공학적 도구로 경사하강법에 관한 시뮬레이션을 하면서 반복 횟수, 학습률 등의 역할을 이해하게 한다. 수업 활동 시 유의점 경사하강법에 관한 시뮬레이션을 하기 위하여 컴퓨터나 스마트폰 이용이 가능한 환경에서 수업을 진행해야 한다. 프로그래밍보다는 경사하강법에 이용되는 수학적 원리를 이해하는데 중점을 둔다. 평가의 주안점 경사하강법 알고리즘을 스스로 생각해 낼 수 있는지 평가한다. 또한 공학적 도구를 적절히 활용하여 경사하강법에 대한 다양한 시뮬레이션을 하고 탐구할 수 있는지 평가한다.
준비물	활동지, 컴퓨터, 태블릿 PC 또는 스마트폰

수업 단계		교수·학습 활동	평가 활동
도입	수업 환경 조성	2~3명을 1개 모둠으로 구성 - 활동지 배분 및 수업 분위기 조성	
수행	모둠 활동	- 손실함수의 최솟값 지점으로 이동하기 위한 방향과 미분계수 사이의 관계 이해하기 - 경사하강법 알고리즘 찾기 - 직접 계산으로 경사하강법 적용해 보기 - 공학적 도구로 경사하강법 시뮬레이션하기 - 시뮬레이션을 통해 학습률에 대해 탐구하기	포트폴리오 평가
정리		학습 내용 정리	



경사하강법

년 월 일



어두운 밤 산을 내려오는 방법은 무엇일까? 다음 물음에 답해 보자.

우리가 등산을 갔다가 불행히도 어두운 밤에 험난한 산속에서 조난을 당했다고 상상해 보자. 어둡기 때문에 무엇도 제대로 보이지 않는다. 지도는 없지만 다행히 손전등이 있는 상황이다. 하지만 이 손전등은 너무나 작고 성능이 떨어져서 지형 전체를 보기는 커녕 불과 몇 미터 앞도 살피기 어려운 수준이다. (신경망 첫걸음 中)

1 이런 상황에서 산을 내려오기 위해 우리는 어떻게 행동해야 할까?

후레쉬로 근처를 확인해 보고 경사가 내려가는 쪽으로 조금씩 계속 움직인다.



경사하강법에 대해 알아보자.

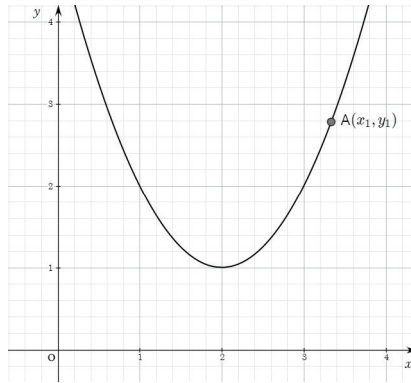
온도와 아이스크림 자료에서 추세선 $y=ax$ 에 대한 손실함수는

$E(a) = \frac{1}{10}(712a^2 - 1566a + 871)$ 과 같이 이차함수 형태로 나타났었다. 이차함수의

최솟값은 계산을 통해 쉽게 구할 수 있으므로 우리는 $a = \frac{783}{712} = 1.0997\cdots$ 일 때 최소가 됨을 수학적으로 구하였다. 그러나 실제 인공지능에서 활용하는 예측 모델은 추세선 $y=ax$ 보다 훨씬 복잡하기 때문에 함수의 형태가 복잡하거나 변수가 2개 이상인 경우가 많다. 따라서 복잡한 함수에서도 최솟값을 찾아갈 수 있는 방법이 필요하다.

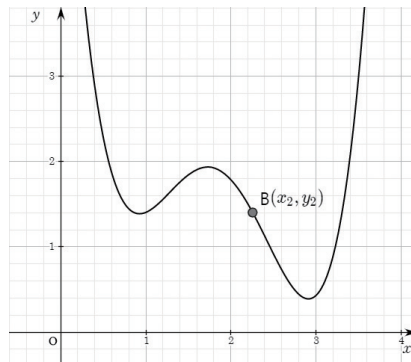
경사하강법은 어떤 a 의 값으로부터 시작하여 $E(a)$ 의 값을 감소시키는 방향으로 a 의 값을 조금씩 수정하는 작업을 아주 많이 반복함으로써 $E(a)$ 가 최소가 되도록 하는 a 의 값을 찾아가는 것이다.

1 손실함수 $f(x)$ 의 그래프 위의 점 $A(x_1, y_1)$ 이 있다. 이 함수가 최소가 되는 지점으로 점 A 를 이동시키려면 A 의 x 좌표를 어느 방향으로 움직여야 할까? 그 방향은 함수 $f(x)$ 의 $x=x_1$ 에서의 미분계수와 어떤 관계가 있는가?



미분계수는 접선의 기울기이므로 양수이다. 함수가 최소가 되는 지점으로 점 A 를 이동시키려면 x 좌표를 왼쪽으로 이동시켜야 한다. 즉, 미분계수의 부호와 반대되는 방향으로 x 좌표를 이동시켜야 한다.

2 손실함수 $f(x)$ 의 그래프 위의 점 $B(x_2, y_2)$ 가 있다. 이 함수가 최소가 되는 지점으로 점 B 를 이동시키려면 B 의 x 좌표를 어느 방향으로 움직여야 할까? 그 방향은 함수 $f(x)$ 의 $x=x_2$ 에서의 미분계수와 어떤 관계가 있는가?



미분계수는 접선의 기울기이므로 음수이다. 함수가 최소가 되는 지점으로 점 B 를 이동시키려면 x 좌표를 오른쪽으로 이동시켜야 한다. 즉, 미분계수의 부호와 반대되는 방향으로 x 좌표를 이동시켜야 한다.

3 1과 2를 참고하여 임의의 x_1 에서 시작하여 충분히 작은 양수 η^9 만큼 100번 움직여서 손실함수 $f(x)$ 가 최소가 되도록 하는 x 의 값을 찾는 알고리즘을 만들어 보자.

- 1) 현재 x 의 값에 대한 미분계수 $f'(x)$ 를 구한다.
- 2) x 의 값을 $-\eta f'(x)$ 만큼 변화시킨다.
- 3) 이와 같은 과정을 100번 반복한다.

4 손실함수를 간단하게 $E(x) = (x-2)^2 + 1$ 로 가정하고 최초의 x 의 값이 4, 학습률 $\eta = 0.3$ 일 때, 함수의 최솟값을 찾기 위해 x 의 값을 4번 수정해 보자.

$E(x) = (x-2)^2 + 1$ 의 x 의 값에 대한 미분계수는 $E'(x) = 2(x-2) = 2x - 4$ 이다.

학습률을 $\eta = 0.3$ 으로 두면, x 의 값을 수정하는 식은 다음과 같다.

$$x - 0.3(2x - 4)$$

이제 $x_1 = 4$ 에서부터 시작해 보면,

$$x_1 = 4$$

$$x_2 = 4 - 0.3(2 \cdot 4 - 4) = 2.8$$

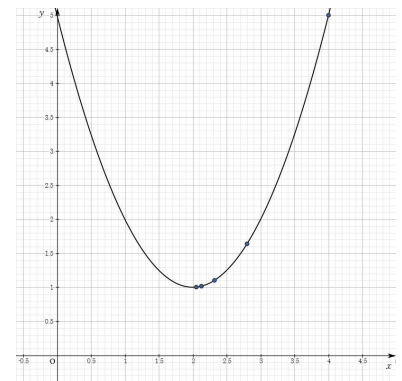
$$x_3 = 2.8 - 0.3(2 \cdot 2.8 - 4) = 2.32$$

$$x_4 = 2.32 - 0.3(2 \cdot 2.32 - 4) = 2.128$$

$$x_5 = 2.128 - 0.3(2 \cdot 2.128 - 4) = 2.0512$$

⋮

조금씩 2에 가까워지고 있음을 확인할 수 있다.



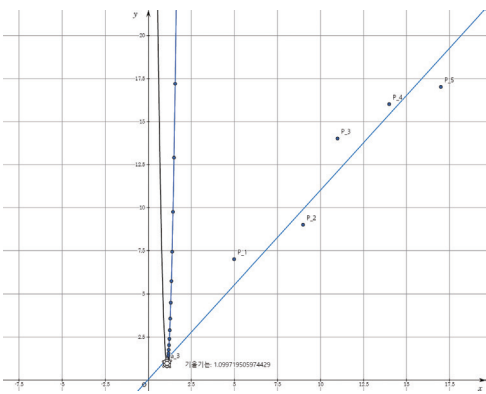
5 아이스크림 문제의 손실함수 $E(a) = \frac{1}{10}(712a^2 - 1566a + 871)$ 가 최소가 되도록 하는 a 의 값을 찾는 알지오매스 코드를 (AI 웹실험실)의 '경사하강법'에서 확인할 수 있다. 다음 물음에 답해 보자.

- 1) (AI 웹실험실)에 접속하여 경사하강법 시뮬레이션을 실행해 보자. 우리는 이차함수의 성질에 의해 $E(a)$ 가 $a = \frac{783}{712}$ 에서 최솟값을 가짐을 이미 알고 있다. 경사하강법을 통해서도 동일한 값을 찾았는가?

1.099719505974429로 매우 근사한 값을 찾았다.

9) η 는 학습률이라고 하는데 x 의 값을 한 번에 얼마만큼 수정하는지를 결정하는 값이다.

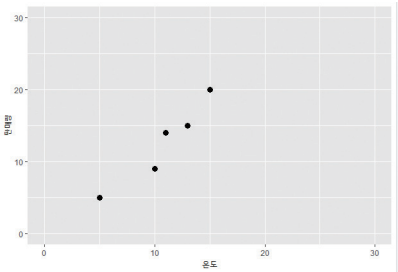
2) 다음은 경사하강법에 대한 시뮬레이션 코드와 실행 결과 화면이다.



‘학습용 자료’, ‘학습률’, ‘초기 a 의 값’, ‘반복 횟수’, ‘ a 의 값을 수정하는 식’을 정하는 코드는 어디인지 찾아서 표시해 보자.

6 단계에 따라 경사하강법을 이용하여 다음 자료를 가장 잘 설명할 수 있는 $y=ax$ 꼴 추세선을 찾고 예측에 활용해 보자.

온도(도)	판매량(개)
5	5
10	9
11	14
13	15
15	20



온도에 따른 아이스크림 판매량

1) 손실함수 $E(a)$ 를 구해 보자.

$$\begin{aligned} E(a) &= \frac{1}{10} \sum_{i=1}^5 (y_i - ax_i)^2 \\ &= \frac{1}{10} \{ (5a-5)^2 + (10a-9)^2 + (11a-14)^2 + (13a-15)^2 + (15a-20)^2 \} \\ &= \frac{1}{10} (640a^2 - 1528a + 927) \end{aligned}$$

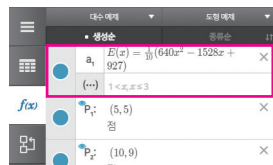
2) 경사하강법에서 a 의 값을 수정하는 식을 구해 보자.

a 를 $a - \eta \frac{1}{5} (640a - 764)$ 로 수정해야 한다.

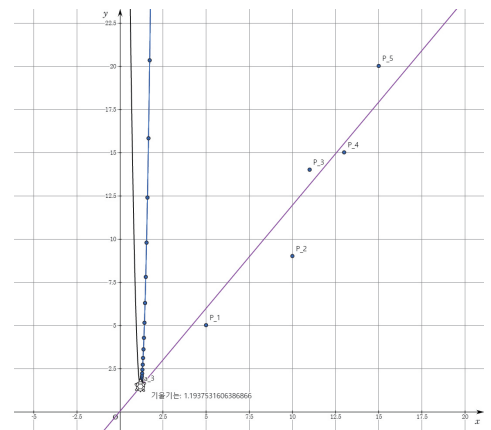


3) 5번 시뮬레이션 코드의 일부를 수정하여 경사하강법을 실행하고 손실함수가 최소가 되도록 하는 a 의 값을 구해 보자. (단, 초기 a 의 값은 4로 한다.)

※ 손실함수는 대수창에서 수정할 수 있다.



다음과 같이 코드를 수정하여 실행할 수 있다.



4) 온도가 30도일 때 아이스크림의 판매량은 몇 개가 될 것으로 예측되는가?

3)의 결과에 의해 추세선을 $y=1.1937x$ 라고 하면, 30도일 때 아이스크림의 판매량은 약 36개로 예측된다.

• **활동팁** 파이썬을 이미 다룰 수 있는 학생들을 대상으로 수업할 때는 다음과 같은 파이썬 코드를 활용 할 수 있다.

```
import numpy as np
import matplotlib.pyplot as plt

# 자료
X = [5, 10, 11, 13, 15]
Y = [5, 9, 14, 15, 20]

# 반복 횟수
iters = 100

# 학습률
alpha = 0.001

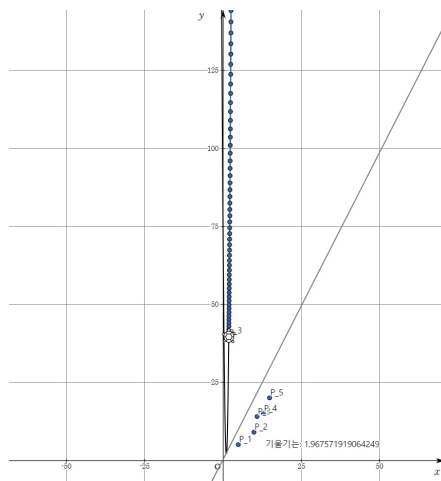
# a의 초기값
a = 4

# a의 값 수정 반복
for k in range(iters):
    a = a - alpha * 1/5 * (640*a - 764)
    if (k % 100 == 0):
        print("반복 = %d a의 값은 %f" % (k, a))

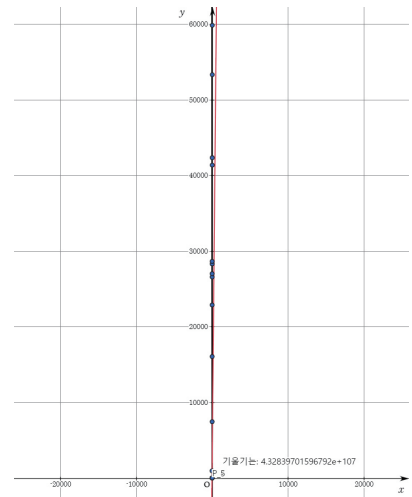
# 그래프
plt.scatter(X, Y)
x = np.arange(0, 20, 0.1)
y = a * x
plt.plot(x, y)
```



7 6번 시뮬레이션에서 학습률을 $\eta=0.0001$ 또는 $\eta=0.1$ 로 두면 어떤 일이 일어나는지 확인 하고 학습률 η 를 어떻게 정해야 할지 생각해 보자.



$\eta=0.0001$ 인 경우 학습률이 너무 작아서 100번의 반복으로도 최솟값을 찾지 못한다. 즉, 학습 속도가 너무 느리다는 단점이 있다.

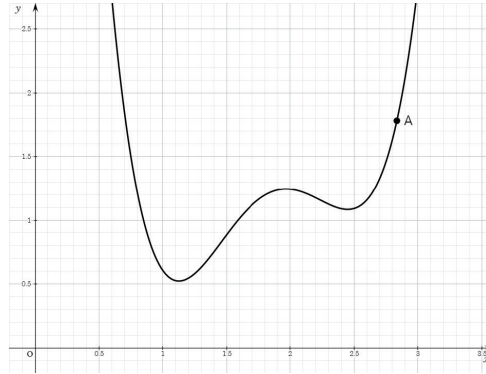


$\eta=0.1$ 인 경우 a 의 값이 무한대로 커진다. 학습률이 너무 커서 최소인 지점을 넘어가면서 발생하는 일 이다.

즉, 너무 작은 학습률은 학습에 지나치게 많은 시간이 걸리게 하거나 학습이 완전히 이루어지지 않게 할 수 있으며, 너무 큰 학습률은 무한대로 발산할 수 있다. 따라서 적절한 값을 사용해야 한다.

8 경사하강법으로 모든 함수의 최솟값을 찾을 수 있을까? 찾을 수 있다면 그 이유를 설명하고, 찾을 수 없다면 예를 하나 들고 설명해 보자.

경사하강법은 접선의 기울기가 0인 지점을 조금씩 이동하면서 찾아가는 과정이며, 접선의 기울기가 0인 지점 근처에 도착하면 움직임이 거의 없어진다. 따라서 다음과 같은 함수에서 점 A의 위치에서부터 경사하강법을 시작했다면 $x=2.5$ 근처에 있는 접선의 기울기가 0인 지점에 도달할 것이며, $x=1.1$ 근처에 있는 실제 최솟값에는 도달하지 못할 것이다. 따라서 함수의 개형에 따라서는 초깃값의 위치에 따라 경사하강법으로 함수의 최솟값을 찾기 어려울 수 있다.



지도의 주안점

- 미분계수의 부호와 손실함수의 최솟값 지점으로 이동하기 위한 방향 사이의 관계를 스스로 이해할 수 있게 한다. 경사하강법을 직접 계산으로 실행해 볼 수 있는 기회를 제공한다. 공학적 도구로 경사하강법에 관한 시뮬레이션을 하면서 반복 횟수, 학습률 등의 역할을 이해하게 한다.

핵심 발문

- 학습률이 너무 작거나 너무 크면 경사하강법은 어떻게 진행될까?



관련 교과 역량

- 추론
 - 미분계수의 부호와 경사하강법의 이동 방향과의 관계를 바르게 추론할 수 있는가?
- 정보 처리
 - 코드를 적절히 수정하여 새로운 상황에 적합한 경사하강법 시뮬레이션을 구성할 수 있는가?

평가 결과 피드백

- 경사하강법의 절차를 생각하지 못하는 경우, 함수의 그래프에서 최솟값을 찾아가는 시각적 방법을 다시 한번 상기시킨다.
- 시뮬레이션 코드 수정에 어려움을 겪는 경우, 기존 코드의 의미를 하나씩 이해할 수 있도록 적절한 발문을 통해 유도한다.

참고 문헌 및 자료 출처

참고 문헌

고선규, 『인공지능과 어떻게 공존할 것인가』, 타커스, 2019.
과학기술정보통신부 웹진, 『인공지능이란?』, 2020.
<http://sejong.korea.ac.kr/user/boardList.do?command=view&boardId=99656&boardSeq=102872>
과학기술정보통신부, 『“IT 강국을 넘어 인공지능 강국으로!” 인공지능 국가전략』, 2019.
교육부(2015), 『수학과 교육과정』, 교육부 고시 제2015-74호[별책 8].
구형일, 『인공지능 및 딥러닝 동향』, 『전기의 세계』, 2018, 67(7), 7-12.
다다사토시, 송교식 역, 『처음 배우는 인공지능』, 한빛미디어, 2017.
박경미 외, 『2015 개정 교육과정에 따른 수학과 교육과정(교수·학습 방법 및 평가) 개발 연구』, 한국과학창의재단, 2015.
아카이시 마사노리, 신상재 역, 『딥러닝을 위한 수학』, 위키북스, 2018.
아키바 신야, 스기야마 아세이, 데라다 마나부, 이종민 역, 『그림으로 공부하는 머신러닝 알고리즘 17-머신러닝 도감』, 제이펍, 2019.
이상구, 이재화, 함윤미, 『인공지능(Artificial Intelligence)과 대학수학 교육』, 『수학교육논문집』, 2020, 34(1), 1-15.
이상구, 함윤미, 이재화, 『뉴노멀(New Normal) 시대 대학수학교육에서의 과정중심 PBL 평가-인공지능을 위한 기초수학』 강좌 사례를 중심으로-, 『수학교육논문집』, 34(4), 421-437.
이시카와 아키히코, 신상재, 이진희 공역, 『인공지능을 위한 수학』, 프리렉, 2018.
히가시나카 류이치로, 진솔 역, 『아무것도 모르고 시작하는 인공지능 첫걸음』, 한빛미디어.
NCTM, 『Curriculum and Evaluation Standards for School Mathematics』, NCTM, 1989.
NCTM, 『Principles and Standards for School Mathematics』, NCTM, 2000.
이원상, 『머신 러닝을 위한 수학 with 파이썬, R』, 길벗, 2020.
이주열, 『인공지능 이미지 인식 기술 동향』, TTA저널, 2020, 44-51.
타리크 라시드, 『신경망 첫걸음』, 한빛미디어, 2017.

자료 출처

35~36쪽 인공지능의 역할 <https://news.samsung.com/kr>
53쪽 순서도 그리기 <https://app.diagrams.net>
61쪽 기계학습 경험하기 <https://teachablemachine.withgoogle.com>
83쪽 기사 https://news.sbs.co.kr/news/endPage.do?news_id=N1006071972&plink=SEARCH&cooper=SBSNEWSSEARCH&plink=COPYPASTE&cooper=SBSNEWSEND

88쪽 다양한 시각화 방법 <https://datavizproject.com/>
88~89쪽 다양한 시각화 방법 <https://developers.google.com/chart>
90~91쪽 워드 클라우드 <http://wordcloud.kr/>
91쪽 기사 <https://news.joins.com/article/23922778>
95쪽 한국어/영어 불용어 예시 <https://www.ranks.nl/stopwords>
103쪽 오토드로우 <https://www.autodraw.com>
105쪽 지도학습과 비지도학습 <https://ellun.tistory.com/103>
108쪽 이미지의 행렬 <http://assets.runemadsen.com/classes/programming-design-systems/pixels/index.html>
110쪽 채널값 https://www.ginifab.com/feeds/pms/color_picker_from_image.php
126쪽 유튜브 영상 화면 <https://www.youtube.com/watch?v=BpB4pYOV25Q>
128쪽 단어 벡터 연산 예시 <https://word2vec.kr/search/>
132쪽 추천 서비스 제공 예시 www.kyobobook.co.kr
134쪽 감성 사전 예시 <https://www.kaggle.com/ratman/sentiment-lexicons-for-81-languages>
134쪽 감성 사전 <http://dilab.kunsan.ac.kr/knu/knu.html>
142쪽 MNIST <http://naver.me/5w54NGG2>
143쪽 숫자 인식 <https://tensorflow-mnist.herokuapp.com>
149쪽 틱택토 <https://www.google.com/search?q=tic-tac-toe>
151쪽 알파고와 확률 <http://tim.hibal.org/blog/alpha-zero-how-and-why-it-works/>
151쪽 바둑과 확률 <https://sites.google.com/site/hkleebrain/home/-site-hkleebrain-home-presentation/alphago>
151쪽 인기 동영상 빅데이터 참고 자료 <https://www.kaggle.com/datasnaek/youtube-new>
151~152쪽 인기 동영상 빅데이터 <https://docs.google.com/spreadsheets/d/1fMDnNGUSNkrG0Ad79CMuCyE6BVYtDoqAMAQN5clj4/edit#gid=451951940>
152쪽 동영상 카테고리 분류 <https://drive.google.com/file/d/1SJ8LDVLU2Uwv3EUJalKOPBxu9GZnAaqq/view>
158쪽 스팸 메일 <https://pctechmag.com/2018/05/spam-mail-101-clear-your-basics-about-this-phishing-problem/>
169~172쪽 알지오매스 <https://www.algeomath.kr/algeomath/app/make.do>
170쪽 페이스북 이용자 수 <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>
171쪽 알지오매스 <http://me2.do/GUGmryzS>
171쪽 인스타그램 이용자 수 <https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/>