

3차 일석이조 조별보고서	
작성일 : 2023년 9월 19일	작성자 : 유정훈
조 모임 일시 : 9월 19일	모임 장소 : 구글미트
참석자 : 이준용, 유정훈, 김동규, 이학빈, 탁성재	조원 : 이준용, 유정훈, 김동규, 이학빈, 탁성재
구분	내용
학습 범위와 내용	1. 3주차 온라인 강의 내용 2. 교재 2장 내용
논의 내용 (조별 모임 전에 조장이 지시)	2.1.1. 벡터와 행렬 2.1.2. 놔와 유사도 2.1.3. 퍼셉트론의 해석 2.1.4. 선형결합과 벡터공간 2.1.5. 역행렬 2.1.6. 행렬 분해
질문 내용 (모임 전 공지된 개별 학습 범위에서 이해된 것과 못한 것)	<p>Q1</p> <p>목요일에 있었던 수업 내용중에 특이값 분해에서 특이값의 부호는 왜 양수일까? 에 대해서 논의해 보았습니다.</p> <p>A1</p> <p>분해(SVD)는 고유값 분해(eigen decomposition)처럼 행렬을 대각화하는 한 방법입니다. 특이값 분해가 유용한 이유는 행렬이 정방행렬이든 아니든 관계없이 모든 $m \times n$ 행렬에 대해 적용이 가능하기 때문입니다.</p> <p>singular value는 AAT, ATA의 고유값분해(EVD)에서 나온 고유값의 square root를 취한 것이라고 했습니다. 그런데, 고유값의 square root는 +, - 2개의 값이 나오는데 왜 +값만 singular value로 잡은 것일까? 사실 원래는 singular value를 +로 잡을수도, -로 잡을수도 있습니다. 하지만 +를 singular value로 잡는 것은 일종의 수학적 약속이라고 할 수 있습니다.</p>

만일 특이값 분해식에서 i 번째 singular value의 부호를 바꿨을 때 U 의 i 번째 열벡터의 부호를 같이 바꾸면 식은 여전히 성립함을 알 수 있습니다. 또한 U 가 직교행렬(orthogonal matrix)란 점도 변하지 않습니다. 그러므로 특이값(singular value)은 항상 0이상입니다.

Q2

분류란 x 라는 데이터가 y 라는 라벨을 가질때, $P(Y|X)$ 을 최대화하는 Y 를 찾는것입니다. 이러한 패턴 인식에서 분류를 학습시킬때, 2가지의 모델을 사용한다고 했습니다. Generative model과 Discriminate model 두 종류입니다.

A2

Generative model은 주어진 트레이닝 셋을 가지고, 해당 데이터의 분포를 따르는 유사 데이터의 생성을 통해 학습을 합니다. 데이터의 분포는 데이터의 갯수가 많으면 많을수록 모양이 확실해지기에 데이터들이 모인 곳을 중심으로 그룹을 만드는데 유리합니다. 즉, 가능성이 높은 부분을 추천해주는 느낌이라고 생각합니다

Discriminate model은 결과까지 주어진 데이터와 해당 데이터로 학습된 패러미터를 직접 이용해 새로운 데이터가 가장 구별 잘 될수있는 구간을 구하는데 중점을 두고 있습니다. 즉, 어느 지점을 넘어가면 데이터의 분류가 달라지는지 해당 구간을 정하는 느낌이라고 생각합니다.

Q3

코사인 유사도란 정확히 무엇인가요?

A3

코사인 유사도란 내적 공간의 두 벡터 간 각도의 코사인 값을 이용하여 측정된 벡터 간의 유사한 정도를 뜻합니다. 각도가 0° 일 때의 코사인값은 1이며, 다른 모든 각도의 코사인값은 1보다 작습니다. 따라서 이 값은 벡터의 크기가 아닌 방향의 유사도를 판단하는 목적으로 사용되며, 두 벡터의 방향이 완전히 같을 경우 1, 90° 의 각을 이룰 경우 0, 180° 로 완전히 반대 방향인 경우 -1의 값을 갖습니다. 이 때 벡터의 크기는

값에 아무런 영향을 미치지 않습니다.

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

코사인 유사도는 어떤 개수의 차원에도 적용이 가능하며 흔히 다차원의 양수 공간에서의 유사도 측정에 자주 이용됩니다. 예를 들어 정보 검색 및 텍스트 마이닝 분야에서, 단어 하나 하나는 각각의 차원을 구성하고 문서는 각 단어가 문서에 나타나는 회수로 표현되는 벡터값을 가집니다. 이러한 다차원 공간에서 코사인 유사도는 두 문서의 유사를 측정하는 매우 유용한 방법입니다.

이렇게 계산된 유사도는 -1 에서 1 까지의 값을 가지며, -1 은 서로 완전히 반대되는 경우, 0 은 서로 독립적인 경우, 1 은 서로 완전히 같은 경우를 의미합니다.

Q4

프로베니우스 놈은 다양한 실용적인 용도가 궁금합니다.

A4

프로베니우스 놈(Frobenius norm)은 행렬의 놈(norm) 중 하나로, 주로 선형 대수와 행렬 분석에서 사용됩니다. 프로베니우스 놈은 다음과 같이 정의됩니다.

만약 행렬 A 가 크기가 $m \times n$ 이라면, 프로베니우스 놈은 다음과 같이 계산됩니다.

$$\|A\|_F = \sqrt{\sum (\sum a_{ij}^2)}$$

여기서 a_{ij} 는 행렬 A 의 i 행과 j 열의 원소이며, \sum 는 해당 원소들에 대한 합을 나타냅니다.

	<p>프로베니우스 놈은 다양한 실용적인 용도가 있습니다. 아래에는 몇 가지 예시를 제공합니다.</p> <ol style="list-style-type: none"> 1. 행렬 유사성 측정: 프로베니우스 놈은 두 행렬 사이의 차이를 측정하는 데 사용됩니다. 두 행렬의 차이를 계산하고, 그 차이가 작을수록 두 행렬은 유사하다고 볼 수 있습니다. 2. 행렬 규제(Regularization): 머신 러닝 및 딥 러닝 모델에서 가중치 행렬의 크기를 제한하고자 할 때 프로베니우스 놈 규제를 사용합니다. 이를 통해 모델의 과적합을 방지하고 일반화 능력을 향상시킬 수 있습니다. 3. 행렬 분해(Matrix Factorization): 프로베니우스 놈은 행렬 분해 기법에서 비용 함수로 사용될 수 있습니다. 예를 들어, 행렬을 낮은 랭크의 두 행렬의 곱으로 근사할 때 비용 함수로 프로베니우스 놈을 최소화하는 방법을 사용할 수 있습니다. 4. 이미지 처리: 이미지 처리에서 특징 맵(feature map)이나 필터(filter)의 크기를 계산하고 조절하는 데에도 프로베니우스 놈이 사용될 수 있습니다. 5. 행렬 분석: 행렬 분석 작업에서 행렬의 크기나 특성을 평가하고 비교하는 데에도 프로베니우스 놈이 유용합니다. <p>프로베니우스 놈은 행렬에 대한 다양한 분석 및 조작 작업에서 중요한 도구로 사용되며, 다른 놈(norm)들과 함께 다양한 응용 분야에서 활용됩니다.</p> <p>Q5</p> <p>퍼셉트론에 대해서 좀 더 자세히 알고 싶습니다.</p> <p>A5</p> <ol style="list-style-type: none"> 1. 인공 신경망의 초기 형태 중 하나로 신경망 및 딥러닝의 기초 개념을 이해하고 발전시킬 수 있는 중요한 개념을 제공했습니다. 딥러닝의 역사는 퍼셉트론에서 시작되었으며, 이후 다양한 신경망 아키텍처가 발전하면서 기계학습 분야가 현재의 발전된 상태에 이르게 되었습니다. 2. 퍼셉트론은 선형 분리 가능한 문제에 대해서만 작동하며, 복잡한 문제해결을 하기엔 어렵다는 점이 있습니다. (XOR 문제와 같이 선형으로 분리할 수 없는 경우)
--	--

	<p>3. 주로 이진 분류(binary classification) 문제를 해결하기 위한 간단한 분류 모델로 사용됩니다. 이진 분류는 여러 응용 분야에서 중요한 문제로, 퍼셉트론은 기본적인 이진 분류 모델로서 유용하게 활용될 수 있습니다.</p> <p>4. 다층 퍼셉트론(MLP)은 여러 개의 퍼셉트론 레이어로 구성된 인공 신경망의 한 형태입니다. 다층 퍼셉트론은 단순한 퍼셉트론보다 더 복잡한 문제를 해결하기 위해 개발되었으며, 여러 개의 은닉(hidden) 층을 포함하고 있어 비선형 문제와 다양한 복잡한 패턴을 학습할 수 있습니다.</p> <p>5. 다층 퍼셉트론(MLP)은 다양한 응용 분야에서 사용되며, 이미지 분류, 자연어 처리, 음성 인식, 게임 인공지능, 금융 예측, 로봇 제어 등 다양한 문제에 적용됩니다. 따라서 범용적으로 유용한 모델입니다.</p>
질문내용	<p>Q: 역행렬의 정리에서 'A의 모든 행이 선형독립이다'와 'A의 모든 열이 선형독립이다'를 묶어서 설명해주시면서 다른 것들의 선형 조합으로 나타나지 않는 것이라고 말씀하셨는데 정확히 이해가 가지 않습니다.</p> <p>Q: 행렬 X1과 행렬 X2의 내적 $X1 \times X2$의 값이 크면 클수록 두 행렬이 유사하다고 하셨는데 왜 그런가요?</p> <p>Q: Generative model 트레이닝 셋이 많아지면 분포도에서 이상치가 더욱 확실히 나타나면서 Discriminate model의 효율에 근접할 것으로 예상되는데 맞는지 궁금합니다. 만약 그렇다면 Generative model이 Discriminate model에 비해서 어떤 장점이 있는지 궁금합니다.</p> <p>$m \times n$ 행렬에 대해 SVD를 할 경우에 보통 $m > n$인 행렬을 대상으로 하는 것이 수학적으로도 해를 결정하는데 적합한 것 같습니다. 만약 $m < n$인 경우에는 연립방정식으로 보면 미지수의 개수가 더 많아서 그냥 생각해봐도 해가 결정되지 않을 것 같습니다.</p> <p>그렇다면 $m < n$인 경우에는 SVD를 하기위해서 어떤 방법을 취해야 하나요?</p>