2주차 조별보고서 (일석이조)

작성일: 2023년 9월 13일	작성자: 이준용
조 모임 일시: 2023년 9월 13일 7교시	모임장소: 구글meet
참석자:201702899 이학빈 201800615 김동규 201802344 유정훈 201803758 탁성재 201904458 이준용	조 원:201702899 이학빈 201800615 김동규 201802344 유정훈 201803758 탁성재 201904458 이준용

구 분	내 용
학습 범위와 내용 (조별 모임 전에 조장이 공지)	1. 간단한 기계 학습의 예 - 선형 회귀 문제 2. 모델 선택 - 과소적합과 과잉적합 , 바이어스와 분산 3. 규제 - 데이터 확대 , 가중치 감쇠 4. 기계 학습 유형 5. 기계 학습의 간략한 역사

논의 내용 (모임 전 공지된 개별 학습 범위에서 이해된 것과 못한 것들)	Q(1) 과잉적합(Overfitting)과 과소적합(Underfitting)은 데이터의 종류 또는 집합에 따라 사용된 모델의 각기 다른 문제점을 보여주는데, 과잉 과소 적합을 해결할 수 있는 방법이 어떤 것들이 있을까? A(1) 과잉적합을 해결하기 위해서는 크게 3가지 방법을 찾았다. 1. 정규화 - 학습데이터의 일부분만 학습시키고, 나머지는 제외해서 테스트에 활용하는 방법이 있다. 2. 훈련 데이터를 많이 모으기 - 모델은 데이터의 양이 적을수록 해당 데이터의 특징 패턴이나 노이즈까지 암기해버려서 Overfitting이 될 확률이 높다. 그래서 데이터의 양을 늘릴수록 모델은 일반적인 패턴을 할습하여 Overfitting을 방지할 수 있다. 3. Dropout - 학습 과정에서 신경망의 일부를 사용하지 않는 방법이다. Dropout 비율을 0.5로 설정한다면 학습 과정마다 랜덤으로 절반의 뉴런만 사용 과소적합을 해결하기 위한 3가지 방법 1. 파라미터가 더 많은 복잡한 모델을 선택하기 2. 모델의 제약을 줄이기(규제 하이퍼파라미터 값 줄이기) 3. 조기종료 시점(overfitting이 되기 전의 시점)까지 충분히 학습하기 Q(2) 최대 우도라는 것은 이미 벌어진 상황에서 해당 상황을 표현하기에 가장 적합한 확률 변수를 구하는 것인데, 기계 학습에서 매우 많은 변수를 가진 케이스의 최대 우도를 구하는 것은 제곱수 이상의 변수를 다룬다는 의미와도 같다. 그렇다면 어떻게 해당 케이스를 다룰 것인가?

	A(2) 책에서의 해결법은 알고리즘을 통한 최적화이다. 최적화 알고리즘 어떻게 하면 우리가 원하는 지점으로 빠르게 이동할 수 있는지에 대한 알고리즘이다. 현재 위치에서 기울기를 찾아내 해당 방향으로 움직이는 경사하강법, 학습률을 조정해 STEP의 크기를 정하는 adagrad, local minimum에 빠지는 것을 방지하기 위해 고안된 모멘텀등 여러 형태의 알고리즘을 조합해 결과를 찾아간다. 해당 알고리즘의 이동에는 “목적함수”라고 불리는 비용함수 또는 손실함수가 더해지며, 해당 목적함수는 이전의 결과값들을 가지고 새로운 샘플의 결과를 예측할 수 있게 도와준다. Q(3) 베이즈 정리는 정확히 무엇이며 쓰이는 예로는 무엇이 있는가? A(3) 기계학습 분야에서 특성들 사이의 독립을 가정하는 베이즈 정리는 나이브 베이즈 분류에서 주로 사용된다. 다른 기계학습 방법론들에 비해 상대적으로 알고리즘이 간단함에도 불구하고 현실세계의 많은 문제를 효과적으로 다룰 수 있다는 장점이 있다. $p(A\|B)=(p(B\|A)p(A))/(p(B))$ p(A)는 A의 사전확률, p(A|B)는 A의 사후확률, p(B|A)는 우도, p(B)는 B의 사전확률을 나타낸다. 쓰이는 예시로는 여러 사람들로부터 측정된 데이터를 바탕으로 성별을 분류해내거나, 문서의 내용에 따라 스팸메일을 구분해낼 수 있다. 이처럼 베이즈 정리는 간단한 디자인과 단순한 가정에도 불구하고 복잡한 실제 상황에 잘 작동된다.

	. Q(4) 가우시안, 베르누이, 이항 분포의 특징은 무엇인가? 그리고 기계학습에서 어떻게 쓰이는가? A(4) 가우시안 분포는 대칭적인 형태를 가지며, 평균을 중심으로 좌우 대칭이다. 이는 분포의 양 끝으로 갈수록 확률이 감소한다는 것을 의미한다. 평균과 표준 편차 두 개의 파라미터로 정의되고 평균은 분포의 중심을 나타내며, 표준 편차는 분포의 폭을 결정한다. 표준 편차가 작을수록 데이터가 평균 주변에 모여 있고, 표준 편차가 클수록 데이터가 분산되어 있다. 베르누이 분포는 주로 두 가지 결과 중 하나가 발생하는 실험의 확률 분포를 모델링하는 데 사용되며 이항 분포 및 다른 분포의 기반 역할을 한다. 이항 분포는 베르누이 실험을 여러 번 반복한 결과를 나타내는 확률 분포이며 주사위 던지기, 동전 던지기, 제품 불량품 검사, 설문조사 결과 분석 등 다양한 응용 분야에서 사용된다. Q(5) 기계 학습에서 선형 회기 알고리즘이 쓰이는 실제 예시가 궁금하다. A(5) 선형 회귀(Linear Regression)는 기계 학습에서 가장 기본적이면서도 널리 사용되는 회귀(Regression) 알고리즘 중 하나이다. 선형 회귀는 입력 변수(또는 특성)와 출력 변수(또는 타겟) 간의 선형 관계를 모델링하는데 사용된다. 다음은 선형 회귀가 사용되는 몇 가지 예시이다: **주택 가격 예측**: 주택 가격을 예측하는 문제에서 주택의 특성(면적, 침실 수, 욕실 수 등)을 기반으로 주택 가격

	을 예측하는 데 선형 회귀가 사용된다. **재무 분석**: 기업의 수익과 지출, 마케팅 비용 등과 같은 여러 요소를 고려하여 기업의 수익을 예측하거나 금융 시장에서 주식 가격 또는 자산 가격을 예측하는 데 선형 회귀를 사용할 수 있다. **의학적 연구**: 환자의 나이, 성별, 혈압, 혈당 농도 등과 같은 요인을 사용하여 특정 질병의 발병 가능성을 예측하는 데 선형 회귀를 활용한다. **제조 및 품질 향상**: 제조 공정에서 센서 데이터를 수집하여 제품의 품질을 예측하거나 제품 생산 속도를 최적화하기 위해 선형 회귀 모델을 사용한다. **자연어 처리**: 텍스트 데이터에서 단어 수, 문장 길이, 문서 내 특정 키워드 등을 사용하여 문서의 난이도를 예측하거나 텍스트의 어떤 특성과 관련된 결과를 예측하는 데 선형 회귀를 활용한다. 이러한 예시 외에도 선형 회귀는 다양한 분야에서 사용되며, 데이터와 관련된 선형 관계를 모델링할 때 유용한 도구 중 하나이다. Q(5-1) 주택 가격 예측에서 선형 회기 모델을 채택하는 이유는 무엇일까? A(5-1) 주택 가격 예측에서 선형 회귀 모델을 채택하는 이유는 다음과 같다. **선형 관계의 가정**: 선형 회귀 모델은 주어진 입력 변수와 출력 변수 간에 선형 관계를 가정한다. 주택 가격에 영향을 미치는 다양한 요인들이 선형적으로 작용할 수 있으며, 이러한 가정이 합리적으로 성립하는 경우가 많다. **간결하고 해석 가능한 모델**: 선형 회귀 모델은 다른 복잡한 모델들에 비해 간결하며 해석하기 쉽다. 주택 가격을 예측하는 데에 있어서도 어떤 변수가 어떤 정도의 영향을 미치는지를 이해하기 쉽다. 예를 들어, 주택의 크기가 1 단위 증가할 때 가격이 얼마나 증가하는지 등을 직관적으로 파악할 수 있다. **데이터가 선형성을 가질 수 있음**: 주택 가격을 예측하는 데 필요한 변수들 중 많은 변수들이 선형 관계를 가질 수 있다. 예를 들어, 주택의 크기가 증가할수록 가격이 증가하는 경향이 있을 수 있다.

	계산 효율성: 선형 회귀 모델은 계산적으로 비교적 간단하며 효율적으로 학습할 수 있다. 이는 대규모 데이터셋에서도 상대적으로 빠른 훈련과 예측을 가능하게 한다. 그러나 주택 가격 예측 문제에는 선형 관계 이외의 복잡한 비선형 요소도 있을 수 있다. 이런 경우에는 다른 고급 기계 학습 알고리즘을 고려할 수 있다. 예를 들어, 결정 트리, 랜덤 포레스트, 뉴럴 네트워크 등의 모델들이 선형 회귀보다 더 복잡한 데이터 패턴을 모델링할 수 있다. 선택한 모델은 데이터와 문제의 특성에 따라 다를 수 있으며, 모델 선택은 주어진 문제에 대한 이해와 실험을 통해 결정되어야 한다. Q(6) 파라미터 조정과정에 있어 상세한 과정이 궁금하다. A(6) 손실 함수를 최소화하기 위해서 가중치와 절편을 조정한다. 이를 위해서 '경사하강법'과 같은 최적화 알고리즘을 설정하고, 파라미터를 반복적으로 업데이트하여 손실 함수를 최소화하는 값을 찾는다. +) 경사하강법이란 경사하강법(Gradient Descent)은 기계 학습과 최적화에서 널리 사용되는 반복적인 최적화 알고리즘 중 하나이다. 이 알고리즘은 함수의 최솟값을 찾거나, 모델의 파라미터를 조정하여 손실 함수를 최소화하는 데 사용되며, 주로 다음과 같은 세 단계로 동작한다. **초기화**: 먼저 최적화하려는 함수나 모델의 파라미터를 초기값으로 설정한다. 이 초기값은 보통 무작위로 선택되거나 일부 규칙에 따라 선택된다. **경사 계산**: 현재 파라미터 값에서의 손실 함수(또는 비용 함수)의 기울기(gradient)를 계산한다. 기울기는 현재 위치에서 손실이 가장 빠르게 감소하는 방향을 나타낸다. 이 방향을 따라 이동하면 손실 함수 값을 줄일 수 있다. **파라미터 업데이트**: 기울기를 사용하여 현재 파라미터 값을 업데이트한다. 업데이트 방향은 경사의 반대 방향으로 이동한다. 이때, 학습률(learning rate)이라고 불리는 하이퍼파라미터를 사용하여 얼마나 큰 보폭으로 이동할

	지를 결정한다. 학습률이 너무 작으면 수렴이 느리게 이루어지고, 너무 크면 수렴이 불안정할 수 있다. **반복**: 위의 단계(경사 계산 및 파라미터 업데이트)를 여러 번 반복한다. 일반적으로 경사 하강법은 손실 함수 값이 충분히 작아질 때까지 또는 미리 정의된 반복 횟수에 도달할 때까지 계속 실행된다. 경사 하강법은 이러한 과정들을 통해 주어진 손실 함수에 대해 최적의 파라미터 값을 찾는 데 사용된다. 이 과정은 함수의 곡선을 따라 점진적으로 내려가는 것과 유사하며, 최솟값에 도달할 때까지 반복된다. 경사 하강법은 다양한 머신러닝 모델의 훈련 및 최적화에 사용되며, 딥러닝과 같은 신경망 모델의 훈련에도 적용된다.
질문 내용	1. 정보이론의 기본원리인 '확률이 작을수록 많은 정보가 있다'라고 했는데 정확하게 어떤 의미인지 잘 이해가 가지 않습니다. 2. 과잉적합과 과대적합에 대해 다같이 논의하던 중 feature라는 용어의 대해 궁금증이 생겼습니다. 과소적합을 해결하기 위해서 데이터를 증가시키는 것보다 feature를 증가시키라는 말이 있었습니다. 이 말의 의미는 모델의 유용한 특징(feature)을 더 많이 보여줄 수 있는 데이터를 증가 시키라는 뜻인지 궁금합니다. 3. 가우시안, 베르누이, 이항 분포 이외에도 자주 쓰이는 분포가 있는지 궁금합니다.
기타	

조 운영 지침

1. 매주 1회 정해진 시간에 구글meet을 이용하여 1시간 정도의 조모임을 갖는다.
2. 각 조원은 학습 범위 내의 교재와 강의 자료를 공부한 후에 이해한 내용과 이해하지 못한 내용을 각각 간단하게 정리하여 개별보고서를 작성한다. (1-2쪽으로 충분함) 작성한 개인 보고서는 모임 전에 모든 조원에게 Notion을 사용하여 전송한다.
3. 그 모임의 회의 진행은 순번을 정하여 돌아가면서 진행하고 해당 순번은 조별모임 한 후에 조별보고서를 작성하여 제출기한 안에 열린 게시판에 게시를 한다.
4. 조별 모임에 참석하지 않는다든지 보고서를 작성하지 않는다든지 혹은 지각 등의 조의 단합을 저해하는 조원은 조원들 스스로 학기 초에 정한 규정에 의하여 처리할 수 있다. (벌금 부과나 조 퇴출 등) 이러한 규정들은 조가 결정된 후에 서로 조별로 협의하여 규정을 만들어 제출하며 규정은 계속 개정할 수도 있다. (규정을 소급적용할 수는 없다.)
5. 조별모임을 원하지 않는 사람이나 퇴출된 학생은 다른 조에 동의를 얻어서 합류하거나 보고서 작업을 혼자 진행한다. (조원의 최대 숫자는 학기 초에 정해진다.)
6. 개인 보고서와 조별 보고서 모두 “자료조사” 혹은 교재 내용을 요약 정리하는 것에 중점이 있는 것이 아니라 자신이 혹은 조원들이 잘 모르겠는 것들 이해되지 않는 것들이 무엇인지를 파악하는 데 중점을 둔다.
7. 작성된 조별 보고서는 수업시간 혹은 과목 홈페이지 게시판을 통하여 설명될 것이다.