

1주차 일석이조 조별보고서	
작성일 : 2023년 9월 6일	작성자 : 탁성재
조 모임 일시 : X	모임 장소 : X
참석자 : 이준용, 유정훈, 김동규, 이학빈, 탁성재	조원 : 이준용, 유정훈, 김동규, 이학빈, 탁성재
구분	내용
학습 범위와 내용	1. 1주차 온라인 강의 내용 2. 교재 1장 내용
논의 내용	*다소 긴박한 제출기한을 사유로, 해당 주차에는 조원 각자 공부하며 가졌던 의문점과 이를 해결하고 알게 된 점들을 공유하는 시간을 가졌습니다.
질문 내용	<p><이준용> Q) 기계 학습의 궁극적인 목표는 오류의 최소화 인데 Loss를 최소화하는 방법에는 어떤 것이 있는가? A) Loss를 최소화하는 방법 중에 하나는 경사하강법이 있다. 경사하강법이란 예측값과 정답값 간의 차이인 손실함수(Loss function)이 최소가 되는 값을 찾는 알고리즘이다. 풀어서 설명하면 학습 데이터 입력을 변경할 수는 없으므로, 손실 함수 값의 변화에 따라 가중치(weight) 혹은 편향(bias)을 업데이트해야 한다. 그럼 최적의 가중치(Weight)를 찾는 법을 수식으로 나타내면 다음과 같다.</p> $w_n = w_{n-1} - \alpha \nabla f(w_{n-1})$ <p>목표 함수인 손실 함수를 비용 함수 w에 대해 편미분하고, 이를 학습률(Learning rate)과 곱한 값을 앞서 선정한 w의 편미분 값에서 빼준다. 위 수식을 통해 손실 함수의 값이 거의 변하지 않을 때까지 가중치를 업데이트하는 과정을 반복한다. 그러면 우리가 목표로 하는 손실 함수 그래프에서 값이 낮은 지점으로 경사를 타고 하강한다고 해서 '경사하강법' 이라 한다. 이외에도 Loss를 최소화하는 방법으로는 배치 경사 하강법, 확률적 경사 하강법, 미니배치 경사 하강법 등이 있다.</p> <p><유정훈> Q) MNIST에 대해 정확한 이해가 되지 않는다. A) MNIST는 손으로 쓴 숫자들로 이뤄진 대형 데이터베이스이며, 다양한 화상 처리 시스템을 트레이닝하기 위해 일반적으로</p>

	<p>사용된다. 이 데이터베이스는 또한 기계 학습 분야의 트레이닝 및 테스트에 널리 사용된다. 총 육 만 개의 트레이닝 이미지와 만 개의 테스트 이미지를 포함한다.</p> <p>Q) Fisher's linear discriminant가 이해되지 않는다.</p> <p>A) Fisher's linear discriminant는 데이터를 여러 변수들의 선형 결합으로 표현하였을 때, 서로 다른 그룹을 잘 구분할 수 있게 해주는 coefficient를 찾는 방법이다. 데이터 집합을 선형 결합에 의하여 직선으로 projection 시켰을 때 서로 다른 class label을 갖고 있는 두 그룹이 직선 위에서 구분이 잘 되어야 한다.</p> <p><김동규></p> <p>Q)특징 공간의 역할과 효율 개선은 어떻게 이루어져야 하는가?</p> <p>A) 특징 공간은 어떤 결과에 영향을 미치는 관측값들을 모은 공간을 말한다. 몇 개의 특징을 가지고 결과를 도출하냐에 따라 해당 특징 공간의 차원이 결정된다. 각각의 특징이 반영하는 가중치는 전부 달라질 수 있기에</p> $y \text{ (결과값)} = w_1(x_1 \text{의 가중치}) * x_1(\text{특징 1}) + w_2(x_2 \text{의 가중치}) * x_2(\text{특징 2}) + \dots$ <p>의 형태를 가지게 된다.</p> <p>하지만 모든 특징을 가지고 계산을 하기에는 차원이 높아질수록 효율이 기하급수적으로 떨어지기에, 학습 알고리즘을 통한 '표현학습'으로 새로운 특징 공간을 확인하거나 적게 영향을 주는 특징의 갯수를 줄이는 '차원 축소'를 통해 효율을 높일 수 있다.</p> <p><이학빈></p> <p>Q)차원이 높아짐에 따라 발생하는 현실적인 문제들에는 어떤 것들이 있는가?</p> <p>A)</p> <ol style="list-style-type: none"> 1. 데이터 밀도의 감소: 차원이 증가할수록 데이터 포인트 간의 거리가 증가하게 된다. 이는 데이터 포인트들 간의 관계를 파악하기 어렵게 만든다. 그렇게 다차원 공간에서 데이터가 희박하게 분포되면, 모델이 효과적으로 패턴을 학습하기 어려워진다. 2. 계산 복잡성 증가: 다차원 데이터에서의 모델 학습과 예측은 계산적으로 매우 비용이 높아진다. 이로 인해 모델 훈련
--	--

	<p>시간이 길어지고 메모리 사용량이 증가하며, 실용적인 머신러닝 모델을 훈련시키기 어려워지는 것이다.</p> <p>3. 과적합 가능성 증가: 다차원 데이터에서 모델은 훈련 데이터에 과적합(Overfitting)되기 쉽다. 다차원 특징 공간에서 모델은 더 많은 파라미터를 가질 수 있으며, 이로 인해 훈련 데이터에 대한 복잡한 패턴을 학습할 수 있지만, 일반화하기 어려워 새로운 데이터에 대한 성능이 떨어질 수 있다.</p> <p>4. 데이터 수집과 관리 어려움: 차원이 증가하면 데이터 수집과 관리가 어려워진다. 다차원 데이터의 수집은 비용이 많이 들 수 있으며, 데이터베이스 및 저장 공간 요구 사항이 높아질 수 있다.</p> <p><특성재></p> <p>Q)기계학습에 있어 인간이 관여하는 부분은 어디인가?</p> <p>A)기계 학습에 있어서 인간의 역할은 여러 가지 단계에서 중요하다. 아래는 기계 학습 프로세스에서 인간의 관여가 필요한 부분들이다:</p> <ol style="list-style-type: none"> 1. 문제 정의: 인간은 기계 학습 프로젝트의 목적을 정의하고 어떤 종류의 모델이 필요한지 결정하는데 관여한다. 예를 들어, 어떤 데이터로부터 어떤 정보를 추출하려는지를 결정하고, 어떤 종류의 예측 또는 분류를 수행할 것인지를 결정한다. 2. 데이터 수집 및 전처리: 인간은 모델을 학습시키기 위한 데이터를 수집하고 정제하는 작업에 관여한다. 데이터의 품질을 평가하고 결측값을 처리하며, 데이터를 모델 학습에 적합한 형태로 가공하는 작업을 수행한다. 3. 특성 엔지니어링: 특성 엔지니어링은 데이터로부터 모델에 입력할 특성(데이터의 속성)을 추출하거나 생성하는 과정이다. 이 단계에서 도메인 전문 지식을 활용하여 어떤 특성이 중요하고 유용한지를 결정하고, 특성을 설계한다. 4. 모델 선택 및 구성: 인간은 어떤 종류의 기계 학습 모델을 사용할 것인지 선택하고, 모델의 구조와 하이퍼파라미터(매개변수)를 조정한다. 이는 모델의 성능에 큰 영향을 미치며, 도메인 지식과 실험을 통해 결정된다. 5. 모델 학습 및 평가: 선택한 모델을 학습시키고, 학습 과정을 관찰하며 모델의 성능을 평가한다. 모델이 얼마나 잘 동작하는지를 평가하고, 필요한 경우 모델을 조정한다. 6. 해석 및 해석: 학습된 모델의 결과를 해석하고, 모델이 어떻게 작동하는지 이해하는 데 인간의 역할이 중요하다. 모델의 예측을 해석하여 실제 의사 결정에 활용할 수 있는 정보를 도출한다. 7. 유지 및 개선: 모델은 환경이 변하면 유지 및 개선되어야 한다. 새로운 데이터를 수집하고 모델을 업데이트하거나, 성능을 향상시키기 위해 모델의 구성을 개선하는 작업을 인간이 수행한다.
--	--

	<p>요약하면, 기계 학습은 인간과 컴퓨터가 협력하여 수행되는 작업이다. 인간은 문제 정의, 데이터 관리, 모델 구성, 결과 해석 등 다양한 단계에서 중요한 역할을 담당하며, 컴퓨터는 데이터 처리 및 모델 학습을 수행한다. 따라서 기계 학습은 인간과 기술의 조화로 이루어지며, 인간의 도메인 전문 지식과 경험이 모델의 성능을 향상시키는 데 중요한 역할을 한다.</p>
기타	<p>금주는 첫주차이므로 개별 보고서를 첨부하지 않았습니다.</p>

참고

조 운영 지침

1. 회의 일정: 조별 회의는 매주 화요일 오후 3 시에 진행되며, 대면으로 약 1 시간 동안 진행된다.
2. 조명: 본 조의 조명은 '일석이조'로 결정하였다.
3. 지각 및 결석에 대한 규정: 지각 또는 결석을 누적 2 회 할 경우, 해당 조원은 기프티콘(아메리카노)을 나머지 조원들에게 제공해야 한다.
4. 보고서 제출 방법: 개별 보고서는 조별 회의 시작 1 시간 전까지 카카오톡 단체 톡방에 공유하도록 한다.
5. 조장 선출 및 순서: 조장은 매주 변경되며, 다음과 같은 순서로 선정됩니다. 탁성재 학우 -> 이준용 학우 -> 유정훈 학우 -> 김동규 학우 -> 이학빈 학우
6. 본 운영 지침은 모든 조원이 공동으로 합의한 사항으로, 조원 간의 원활한 소통과 협력을 위해 준수할 필요가 있다.
7. 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.)
8. 조별모임을 원하지 않는 사람이나 퇴출된 학생은 다른 조에 동의를 얻어서 합류하거나 보고서 작업을 혼자 진행한다. (조원의 최대 숫자는 학기 초에 정해진다.)
9. 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.9. 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.
10. 작성된 조별 보고서는 수업시간 혹은 과목 홈페이지 게시판을 통하여 설명될 것이다.