

1주차 조별보고서 (Default)

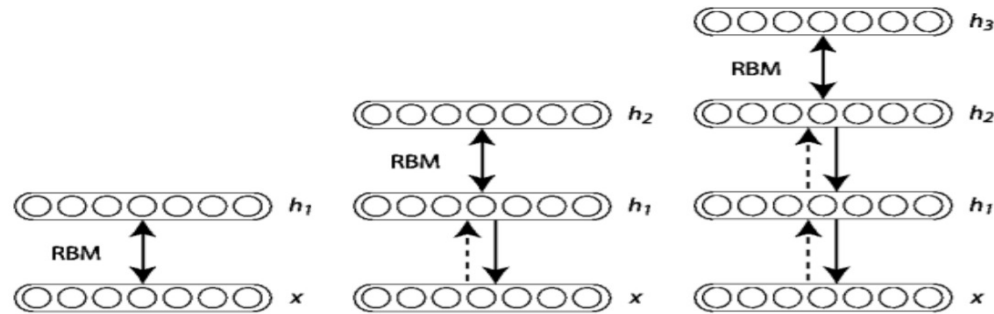
작성일: 2019년 9월 14일		작성자: 위성조	
조 모임 일시: 2019년 9월 14일 오후 10시		모임장소: 추석연휴로 인해 부득이하게 카카오톡으로 진행하였습니다.	
참석자: 위성조, 이충현, 최진성, 김영연, 이재은		조원: 위성조, 이충현, 최진성, 김영연, 이재은	
구	분	내 용	
학습 범위와 내용		1.1 기계 학습이란	
		1.2 특징공간에 대한 이해	
논의 내용		개념 파악 후 배웠던 내용을 복습 및 정리하면서 각자 궁금했던 점의 질의응답	
		Q1. 딥 러닝에서 은닉층을 가진 신경망을 이용한다고 자료에 써있는데, 이 은닉층의 개수와 각 은닉층의 노드의 개수는 어떤 방법을 사용하여 결정 하는지 궁금합니다. 예를 들어, 은닉층이 없는 단층 퍼셉트론에서 해결이 불가능한 XOR문제 등(선형분리 불가한 데이터)은 은닉층을 추가하여 해결 할 수 있지만, 은닉층을 추가하는 것은 차원을 증가시킨다는 것과 동일한 의미인 것으로 판단되는데, 그렇다면 수업자료에 나온 차원의 저주 문제 때문이라도 층을 무한히, 혹은 너무 많이 추가시키는 것은 좋지 않을 것 같습니다. 인터넷으로 찾아보아도 층이 너무 많아지면 overfitting 등의 문제가 발생 할 수 있다는 정보만 찾을 수 있어, 혹시 은닉층의 개수와 노드의 개수를 정하는 방법이 있는지 궁금합니다.	
		A1. Hidden Layer가 증가하게 된다면 기존에 분류할 수 없었던 Data를 구분하는게 가능하다 하지만 지나치게 Hidden Layer를 증가시키면 오히려 전체 Data 분리에 있어서 성능이 하락하게 되는데, 이를 경사 감소 소멸(Gradient Descent Vanishing)이라고 한다. 이 문제를 해결하기 위해선 이전의 RBM(Restricted Boltzmann Machines) 방식 뿐만이 아닌 DBN(Deep Belief Network)이라는 방법을 사용해야 한다고 한다.	

이는 RBM을 이용해서 Multi-Layer Perceptron의 가중치를 들어온 데이터들만을 보고 Unsupervised Learning 방법을 이용하여 학습시켜 이후 다른 알고리즘들을 이용해 가중치를 미세 조정하는 방법이다.

첫 번째 층의 RBM의 사전 훈련이 완료되면 첫 번째 층의 가중치와 데이터를 고정한 채 그 데이터를 입력으로 하여 두 번째 RBM 층을 사전 훈련한다.

이렇게 원하는 층수만큼 RBM을 쌓아 올려 DBN을 완성한다.

이 DBN의 장점은 입력값으로만 학습을 수행한다는 점과, 사용할 데이터의 양이 적을 때 주로 사용할 수 있다는 점이다.

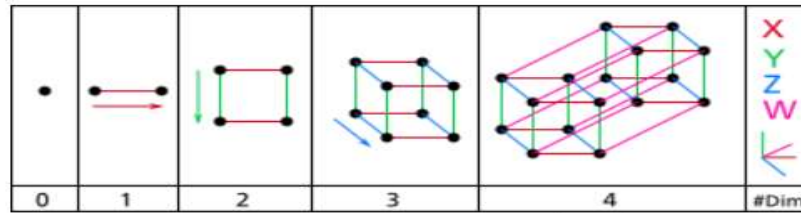


Q2.

수업 시간에 차원과 관련된 설명 중에 “정육면체 안에 구를 넣고, 정육면체의 크기를 무한히 증가시키면 구안에서의 정보들이 없어진다.” 라는 예시를 들었습니다. 이 부분에 대하여 차원의 저주(curse of dimensionality)와 연관시켜 보다 깊이 이해하고 싶습니다.

A2.

기하학적으로 본다면 0차원은 점의 세계, 1차원은 선의 세계, 2차원은 면의 세계, 3차원은 공간의 세계라고 볼 수 있다. 여기서 차원이란 수학에서 공간 내에 있는 점 등의 위치를 나타내기 위해 필요한 축의 개수를 말한다. 그냥 변수의 수라고 이해하면 쉽다.



〈그림 1〉 출처: 위키피디아(차원)

위 그림에서 보듯이, 변수의 수가 늘어나면 차원이 늘고 데이터의 공간이 커진다는 뜻이다. 차원이 커지면서 발생하는 문제를 차원의 저주라고 한다. 차원의 저주는 데이터의 차원이 증가할수록 해당 공간의 크기가 기하급수적으로 증가하기 때문에 동일한 개수의 데이터의 밀도는 차원이 증가할수록 희박 해진다. 위 질문에서 '정육면체의 크기를 무한히 증가시키면 구안에서의 정보들이 없어진다'라는 말은 차원이 증가할수록 정보의 밀도가 감소한다는 뜻으로 해석할 수 있다. 즉, 정보의 '질'을 담당하는 차원은 한정되어 있는데 데이터의 개수와 차원이 무수히 증가하게 된다면, 학습모델의 성능은 떨어질 것이다.

Q3.

기계 학습의 종류 중에는 Supervised Learning과 Unsupervised Learning 외에도 Reinforcement Learning과 Recommender Systems가 있다고 한다. 그렇다면 Reinforcement Learning과 Recommender Systems는 각각 어떤 방법을 이용하여 기계 학습을 시행하는지, 그 외에 다른 방법으로 시행하는 기계 학습은 무엇이 있는지 알고자 한다.

A3.

강화 학습(Reinforcement Learning)은 어떤 환경 안에서 정의된 주체(agent)가 현재의 상태(state)를 관찰하여 선택할 수 있는 행동(action)들 중에서, 가장 최대의 보상(reward)을 가져오는 행동이 무엇인지를 학습하는 것이다. 주체가 환경으로부터 보상을 받는 부분이, 입력과 결과값(label)을 이용한 학습 방법인 지도 학습과 유사해 보이지만 사람으로부터 학습을 받는 것이 아니라 변화되는 환경에서 학습을 한다는 것에서 차이가 있다.

추천 시스템(Recommender system)은 사용자가 생각하는 순위나 선호도를 예측해서 최적의 아이템을 필터링하는 일종의 정보 검색 시스템이다. 추천 시스템은 크게 아이템의 내용을 분석하여 추천하는 내용기반 접근 방식과, 사용자들의 행동 패턴을

분석하여 추천하는 협업적 접근 방식으로 나뉘는데 두가지를 같이 사용하기도 한다.
이 외에도 준지도 학습(Semi-Supervised Learning)방식이 있다. 이는 라벨링이 되어 있거나 또는 되어있지 않은 정보들을 이용하여 더 좋은 학습 모델을 만드는 방식이다.

Q4.

기계학습의 궁극적인 목표는 'Loss의 최소화'인데 어떻게 Loss를 최소화시킬까?

A4.

Loss를 최소화하기 위한 방법에는 경사하강법이 있다.

경사하강법이란 쉽게 말해 Cost function의 기울기가 최소가 되는 값을 찾는 알고리즘으로, 임의의 한점에서 기울기를 구한 다음, 기울어진 방향으로 계속 이동하여 cost function이 최소가 되는 최적의 parameter를 찾으면 된다.

이 과정을 수식화하면 다음과 같다.

$$\theta_{step+1} = \theta_{step} - pg$$

수식을 풀어 말하자면, 기존 세타 값에서 경사를 따라 학습률(p) * 경사값(g)만큼 내려가는 것을 나타낸다. 여기서 경사 값은 방향을 나타내고 학습률은 얼마나 빨리 내려가는지를 정하는 값이다.

다시 말하면 경사 값(g)에 -1을 곱한 방향으로 학습률(p)만큼 움직이는 것

이 과정은 손실 함수의 출력 값이 많이 줄어들지 않을 때까지 반복


(손실 함수의 출력 값이 많이 줄어들지 않았다 = 경사를 거의 다 내려왔다.)


Q5.

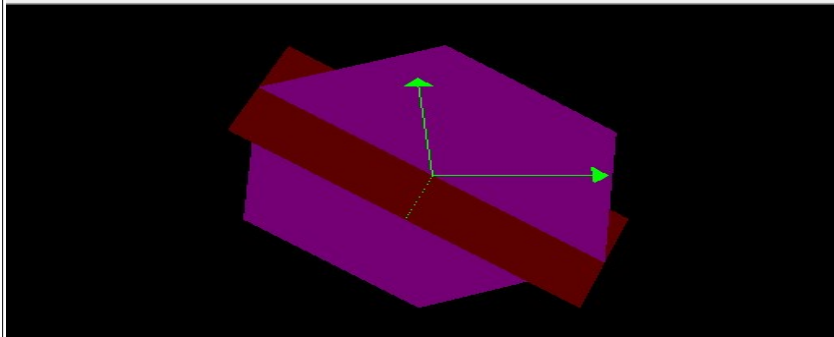
수업 중 고차원을 직선의 방정식으로 표현하는 것에 대하여 얘기가 나왔고 그 중 3차원 공간의 2차원 평면을 의미하는 식은 $ax + by + cz + d = 0$ 이라고 하였다. 그렇다면 4차원 공간에서 3차원을 의미하는 식은 $ax + by + cz + dw + e = 0$ 으로 나타내면 되는 것인가? 만약 그렇게 된다면 이유는 무엇인가?

A5.

2차원에서 직선을 의미하는 식은 $ax + by + c = 0$ (a, b, c 는 상수) 인데, x 값 혹은 y 값이 변화할 경우 각 축의 방향으로 얼마만큼 이동해야 하는지의 정보를 주기 위해 x 와 y 변수를 사용하였다. 따라서 3차원에서 z 축의 변화 정보를 주기 위해 z 변수를 추가하였으며, $ax + by + cz + d = 0$ 의 형태가 된다.

 $x + y + z = 0$

 $x + y = 0$



※ 빨간색 평면의 경우 z 축에 대한 정보가 없기에, xy 평면에서의 그래프를 그대로 3차원으로 확장한 모양을 갖는다.

따라서 4차원에서도 각 평면의 변화정보를 주기 위해서 $ax + by + cz + dw + e = 0$ 의 형태가 될 것이라고 유추할 수 있다.

질문 내용	
기타	

조 운영 지침

1. 매주 1회 정해진 시간과 장소에 모여서 1시간 정도의 조모임을 갖는다.
2. 조장은 모임 전에 학습할 범위를 조원들에게 통보한다.
3. 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 전송한다.
4. 그 모임의 회의 진행은 성조-영연-진성-재은-충현 순으로 돌아가면서 진행하고 해당 순번은 조별모임 한 후에 조별보고서를 작성하여 다음 수업 시간 전에 과목 웹 페이지에 게시를 한다.
5. 조별 모임에 참석하지 않는다든지 보고서를 작성하지 않는다든지 혹은 지각 등의 조의 단합을 저해하는 조원은 조원들 스스로 학기 초에 정한 규정에 어긋난 행동을 3번 할 시에는 퇴출한다. (단, 부득이한 일이 있을 경우, 사전에 조원들에게 알린다.)
6. 조별모임을 원하지 않는 사람이나 퇴출된 학생은 다른 조에 동의를 얻어서 합류하거나 보고서 작업을 혼자 진행한다. (조원의 최대 숫자는 학기 초에 정해진다.)
7. 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.
8. 작성된 조별 보고서는 수업시간 혹은 과목 홈페이지 게시판을 통하여 설명될 것이다.

첨부 개인 레포트

이름	위성조	학번	201402033
구분	내용		
학습 범위	기계 학습이란 특징공간에 대한 이해		
학습 내용	<p>기계 학습의 현대적 정의</p> <p>- 어떤 컴퓨터 프로그램이 T라는 작업을 수행한다. 이 프로그램의 성능을 P라는 척도로 평가했을 때 경험 E를 통해 성능이 개선된다면 이 프로그램은 학습을 한다고 말할 수 있다.</p> <p>[개인적 요약 - 피드백을 통해 프로그램의 성능이 개선되면 학습을 한다고 볼 수 있다.]</p> <p>인공지능의 제작 방법</p> <p>초창기 - 지식기반 (사람이 생각하는 물체의 특징을 프로그램으로 기술 / 문제점 - 단추와 8을 구분 못함)</p> <p>현재 - 데이터 기반 (데이터를 넣어 주고 컴퓨터가 스스로 특징을 찾아냄 / 문제점 - 컴퓨터가 어떤 특징으로 물체를 구분하는지 알 수 없음)</p> <p>기계학습은 주어진 데이터들을 바탕으로 새로운 데이터가 들어왔을 때의 결과값을 가장 정확하게 예측할 수 있는 수식 혹은 수식의 매개변수를 찾는 작업.</p> <p>처음부터 최적값을 알 수 없기에, 임의의 값에서 시작하여 점차적으로 성능을 개선한다.</p> <p>기계 학습의 궁극적인 목표는 훈련집합에 없는 새로운 샘플에 대한 오류를 최소화 하는 것이며, 테스트 집합에 대한 높은 성능을 일반화(generalization)이라 부른다.</p> <p>d-차원 데이터</p>		

	<p>특징 벡터 표기 :</p> <p>d-차원 데이터를 위한 학습 모델</p> <p>직선 모델을 사용하는 경우 매개변수 수 = d + 1</p> $y = w_1x_1 + w_2x_2 + \cdots + w_dx_d + b$ <p>2차 곡선 모델을 사용하면 매개변수 수가 크게 증가 / 매개변수 수 =</p> $y = w_1x_1^2 + w_2x_2^2 + \cdots + w_dx_d^2 + w_{d+1}x_1x_2 + \cdots + w_{d^2}x_{d-1}x_d + w_{d^2+1}x_1$ $+ \cdots + w_{d^2+d}x_d + b$ <p>표현 학습(representation learning)</p> <p>좋은 특징 공간을 자동으로 찾는 작업 - 선형 분리 불가능한 원래 특징공간을 변환하여 정확률을 올릴 수 있다.</p> <p>딥 러닝은 다수의 은닉층을 가진 신경망을 이용하여 계층적인 특징 공간을 찾아내며, 왼쪽에서 오른쪽으로 갈수록 고급 특징을 추출함</p> <p>수식은 차원에 무관하게 적용 가능 - 따라서 2~3차원의 저차원에서 식을 고안한 후 고차원으로 확장 적용이 일반적</p> <p>차원의 저주 - 어떤 데이터에 대하여 차원을 충분한 고민 없이 증가시킬 경우, 데이터가 위치할 수 있는 공간이 기하 급수적으로 늘어나, 데이터 분류에 어려움을 겪을 수 있다.</p>
질문 내용	<p>딥 러닝에서 은닉층을 가진 신경망을 이용한다고 자료에 써 있는데, 이 은닉층의 개수와 각 은닉층의 노드의 개수는 어떤 방법을 사용하여 결정 하는지 궁금합니다.</p> <p>예를 들어, 은닉층이 없는 단층 퍼셉트론에서 해결이 불가능한 XOR문제 등(선형분리 불가한 데이터)은 은닉층을 추가하여 해결 할 수 있지만, 은닉층을 추가하는 것은 차원을 증가시킨다는 것과 동일한 의미인 것으로 판단되는데, 그렇다면 수업자료에 나온 차원의 저주 문제 때문이라도 층을 무한히, 혹은 너무 많이 추가시키는 것은 좋지 않을 것 같습니다. 인터넷으로 찾아보아도 층이 너무 많아지면 overfitting 등의 문제가 발생 할 수 있다는 정보만 찾을 수 있어, 혹시 은닉층의 개수와 노드의 개수를 정하는 방법이 있는지 궁금합니다.</p>

이름	이충현	학번	201402665
구분	내용		
학습 범위	기계 학습이란 특징공간에 대한 이해		
학습 내용	<p><머신러닝의 개념></p> <p>크게 두 가지로 나뉜다.</p> <p>Supervised Learning -> 직역하면 지도된 학습이다. 레이블링을 통한 학습(Labeled). Input으로 데이터를 주면 Output으로 학습된 결과물을 보여준다. 따라서 학습을 하기위한 Training Dataset가 필요하다.</p> <p>대표적으로 3가지가 있는데</p> <ul style="list-style-type: none"> -Regression ex)0-100 사이에서 점수예측 -Binary Classification ex)시험에서 P/F예측 -Multi-Label Classification ex)학점 등급(A-F)예측 <p>Unsupervised Learning ->말 그대로 지도되지 않은 학습. 레이블링 없이 데이터를 가까운 기준으로 묶을 때 사용된다. 딱히 답은 없다. 그래서 Input만 주고 Output은 그 상황이나 환경에 반영돼서 나온다.</p> <p>Ex) Word Clustering, Google News Grouping</p> <p>데이터 모델링</p>		

Hypothesis (Linear)

$$H(x) = Wx + b$$

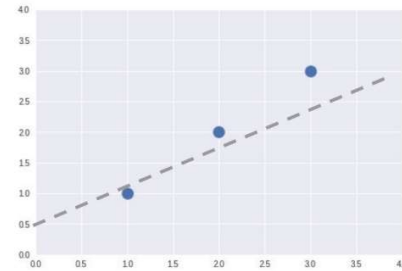


Figure 1

가장 simple하게 직선으로 모델링하자. 수 많은 데이터 중에 $H(x)$ 는 데이터의 평균화를 나타낸다.
따라서 이는 $H(x)$ 라는 가설(Hypothesis)을 세우고 W 는 기울기, b 는 절편이라는 전제하에 표현된다.

Cost

How **fit** the line to our (training) data

$$\frac{(H(x_1) - y_1)^2 + (H(x_2) - y_2)^2 + (H(x_3) - y_3)^2}{3}$$

$$cost(W) = \frac{1}{m} \sum_{i=1}^m (Wx_i - y_i)^2$$

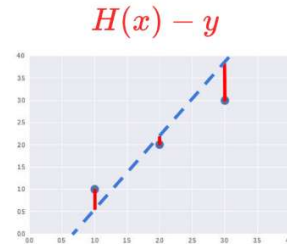


Figure 2

	<p>여기서 ‘학습이 잘 되었다’=’Error가 적다’=’Loss(Cost)가 적다’=’y와 데이터(점)간의 거리가 짧다’ 라고 볼 수 있다. 학습의 궁극적인 목표는 Loss를 최소화 시키기 위함이다. Figure2에서 직선상 거리가 음수가 나올 수 있는데 음수인 거리는 존재하지 않으므로 제곱을 시키고 평균값을 구한다.</p> <p>이렇게 학습이 마치면 예측에 사용하게 된다. 그리고 그 예측과 실제 값의 차이가 최소가 되면 잘 학습되었다고 본다.</p> <p><특징 공간과 공간변환></p> <p>1,2차원 특징 공간</p> <p>1차원 특징 공간은 단순 점으로 표시한다. 2차원 특징 공간은 특정 벡터를 사용한다.</p> <p>다차원 특징 공간</p> <p>직선모델(Hyperplane)을 사용하는 경우의 식</p> $y = w_1x_1 + w_2x_2 + \dots + w_dx_d + b$ <p>고차원의 특이성에서 특정 변수가 N개이면 N차원의 특징 공간을 가진다. 그러나 3차원 이상일때 고차원으로 갈수록 전체 공간에서 데이터가 차지하는 공간이 매우 미비해진다. 새로운 샘플과 훈련 샘플이 멀리 떨어져 있을 가능성이 높다는 뜻이므로 예측을 위해 많은 작업을 하고 저차원일 때보다 예측이 불안정해진다. ->’차원의 저주’</p> <p>해결책으로는 두 가지가 있는데 이론적으로 차원의 크기가 그대로라면 차라리 데이터의 크기를 키워서 밀도를 높이는 방법이 있다. 그러나 데이터의 크기가 무한정 있지는 않으므로 패스. 따라서 어쩔 수 없이 차원 자체를 축소해야 하는 방법이 있다.</p>
질문 내용	Figure2에서 Loss를 구하는 것까지는 이해했다. 근데 문득 드는 게, 기계학습의 궁극적인 목표는 ‘Loss를 최소화’ 시키는 것인데...어떻게 Loss를 최소화 시킬까??

이름	최진성	학번	201403474
구분	내용		
학습 범위	기계 학습이란 특징공간에 대한 이해		
학습 내용	<p>머신러닝이란?</p> <p>Arthur Samuel에 의해 창안되고 Tom Mitchell에 의해 현대의 머신러닝의 개념이 정립됨. (A : 어떠한 작업에 대해서 프로그래밍을 하지 않아도 인간처럼 스스로 배우는 능력이 발현하는 것.) (T : 어떤 작업(Task)을 해서 그 경험(Experience)으로 인하여 성능(Performance)이 증가하였을 경우 기계적 학습을 하였다고 볼 수 있다.)</p> <p>Supervised Learning -> 레이블링을 통한 학습. 데이터 집합을 제공할 경우 그 데이터에 정답이 있는 경우를 의미함. 주로 회귀(Regression)와 분류(Classification) 문제의 해결에 많이 쓰임. (회귀 : 집의 평수 당 가격과 같이 스칼라 값을 연속으로 가진 데이터의 결과를 예측해야 하는 문제.) (분류 : 암세포 검사에서 음성 / 양성, 혹은 어떤 종류의 암세포인지 확인하는 등 결과가 불연속적인 값들로 분류되는 문제.) -Binary, Multi Classification으로 나뉨</p> <p>Unsupervised Learning -> 정답이 되는 데이터가 아닌, 레이블링 없이 데이터를 가까운 기준으로 묶을 때 사용. (= Clustering Algorithm) Ex : SNS에서의 관계 분류, 사람의 음성 분류(Cocktail Party Problem), 기호 분류</p> <p>그 외 : Reinforcement Learning, Recommender Systems..</p>		

초기의 인공지능은 지식기반 방식이 주류였다.

Ex : 단추는 가운데에 구멍이 몇 개 있는 동그란 물체이다.

곧 지식기반 방식의 큰 한계점을 발견하게 됨.(가운데에 구멍이 몇 개 있는 동그란 물체는 모두 단추인가?)

지식기반의 접근에서 데이터 중심이 되는 기계 학습 접근 방식으로 변경

기계학습은 주로 훈련집합을 통한 예측을 주로 시행 -> 가장 정확하게 예측할 수 있는, 즉 최적의 매개변수를 찾는 작업
즉, 기계학습의 초기와 후기의 성능은 상당히 다르다(후기는 데이터가 쌓여 결과값에 대한 어느 정도의 신뢰성을 확보함)
->일반화(테스트 집합에 대한 높은 성능을 의미)

다차원 특징 공간

1~4차원 + Wine의 13차원, MNIST의 784차원, Farm ads의 54877차원 등..

d-차원 데이터의 기본 학습 모델

초평면(Hyperplane)을 사용하는 경우의 식

$$y = w_1x_1 + w_2x_2 + \dots + w_dx_d + b$$

선형 분리 불가능한 특징 공간

선형분리란? 2개 이상의 집합(초평면)을 직선으로 분리하는 것

최대 마진 평면을 통해 집합을 분리할 수 있음.

하지만 집합의 개수가 많아질 경우 선형 분리가 불가능한 공간이 발생할 수 있다.

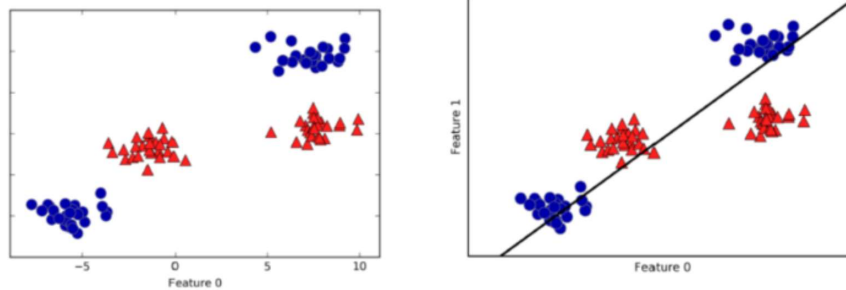
Ex : XOR 집합

OX

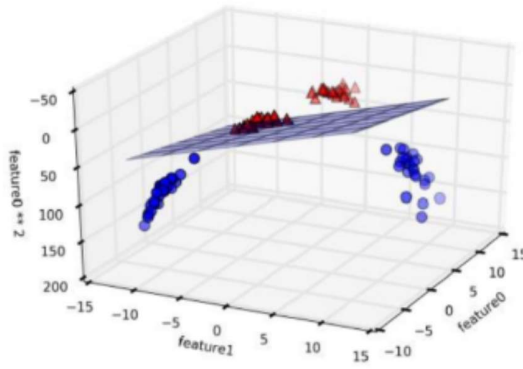
XO

	<p>위의 경우엔 하나의 직선으로 2개의 집합을 분리할 수 없다. 이 경우에는 기존의 특징공간에 차원을 추가하는 방법을 통해 선형 분리를 하는 방법으로 진행해야 한다.</p> <p>Ex : 2차원의 초평면을 3차원으로 발전시켜 초공간을 구성하고, 직선을 초평면 태로 만들어 선형 분리 진행</p> <p>차원의 특이성</p> <p>길이가 1인 선분으로 이루어진 2차원 정사각형과 지름이 1이며 사각형에 내접하는 원이 있다고 하자. 이 경우 가장 먼 꼭짓점 사이의 거리는 이다.</p> <p>차원을 하나 늘려서 길이 1인 선분으로 이루어진 3차원 정육면체에 내접하는 지름 1의 구가 있다고 하자. 이 경우 가장 먼 꼭짓점 사이의 거리는 이다.</p> <p>즉, 차원을 n개 높일수록 꼭짓점 사이의 거리는 이 되어 점점 커지지만, 그에 내접하는 원을 기반으로 하는 물체의 지름은 1이 계속 유지 된다. 즉, 외부의 크기는 점점 커지지만 내부의 크기는 상대적으로 계속 작아지는 현상이 발생한다.</p>
질문 내용	<p>기계 학습의 종류 중에는 Supervised Learning과 Unsupervised Learning 외에도 Reinforcement Learning과 Recommender Systems가 있다고 한다.</p> <p>그렇다면 Reinforcement Learning과 Recommender Systems는 각각 어떤 방법을 이용하여 기계 학습을 시행하는지, 그 외에 다른 방법으로 시행하는 기계 학습은 무엇이 있는지 알고자 한다.</p>

이름	김영연	학번	201500629
구분	내용		
학습 범위	기계학습의 정의와 개념 특징공간과 공간 변환을 소개		
학습 내용	<p>-Supervised learning: 미리 정의된 것을 기준으로 머신러닝 알고리즘을 모델링하고 그 결과 모델을 이용하여 정답을 유추하는 것 à 예측문제</p> <p>1) 회귀(regression) 문제: 연속적이며 기대되는 특정 결과 값의 수치를 예측할 수 있는 경우</p> <p>2) binary classification 문제: 연속성의 띄지 않고 두가지로 결과 값을 분류하는 경우(예: 참/거짓, 성적 평가방식의 pass/fail)</p> <p>3) multi classification 문제: binary classification 처럼 두가지로 결과 값을 분리하는 것이 아닌 그 이상의 결과 값으로 분류하는 경우(예: 성적평가 방식에서 A/B/C/D/F)</p> <p>-Unsupervised learning: 비슷한 것들을 묶어주는 것</p> <p>-d-차원 데이터</p> <p>-선형 분리 불가능한 특징공간</p> <p>à 2차원에서 ‘선형분리’란 두 클래스를 나누는 직선을 찾는 것이다. 두 클래스 사이를 가장 멀리 분리하는 최대 마진 초평면을 찾아야 데이터에 대해 가장 일반화를 잘할 수 있다.</p> <p>그러나 데이터를 2차원 공간에서 표현하였을 때, 선형적으로 분리할 수 있는 경우와 분리하지 못하는 경우가 존재한다.</p> <p>선형적으로 분리할 수 있는 경우에는 최대 마진평면을 찾으면 된다.</p> <p>선형적으로 분리하지 못하는 경우에는 비선형 공간을 위한 kernel을 사용하여, 분리할 수 있는 수준으로 차원을 올려줘야 한다.</p> <p>위의 내용을 좀 더 자세히 이해하기 위하여 그림을 통해 살펴보자.</p> <p>다음은 선형적으로 분리할 수 없는 경우를 나타내는 그림이다.</p>		



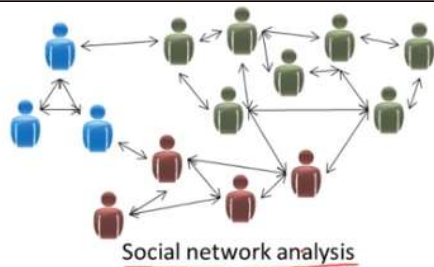
이를 보다 잘 분리하기 위해서는 2차원을 kernel을 이용하여 다음과 같이 고차원으로 분리를 한다면 비선형 관계를 선형으로 나타낼 수 있다.



-차원의 저주

질문 내용 수업 중 고차원을 직선의 방정식으로 표현하는 것에 대하여 얘기가 나왔고 그 중 3차원 공간의 2차원 평면을 의미하는 식은 $ax + by + cz + d = 0$ 이라고 하였다. 그렇다면 4차원 공간에서 3차원을 의미하는 식은 $ax + by + cz + dw + e = 0$ 으로 나타내면 되는 것인가? 만약 그렇게 된다면 이유는 무엇인가?

이름	이재은	학번	201502469
구분	내용		
학습 범위	1.1절: 기계 학습의 역사, 정의와 개념 1.2절: 특징 공간과 공간 변환이란?		
학습 내용	<p>기계 학습의 정의: “경험의 결과로 나타나는, 비교적 지속적인 행동의 변화나 그 잠재력의 변화. 또는 지식을 습득하는 과정 [국립국어원2017]”</p> <p>한마디로, 어떤 분야에서 쌓인 데이터를 토대로 학습을 통해 보다 좋은 성능으로 문제를 해결하는 프로그램을 만드는 것을 뜻한다.</p> <p>-수업 중 논의했던 문제:</p> <p>Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?</p> <p>기계 학습의 알고리즘</p> <p>-지도 학습(Supervised learning): 입력과 결과값(label)을 이용한 학습 분류(classification), 회귀(regression) 등 여러가지 방법에 쓰인다. Ex) 구체적인 학습 모델-> SVM(Support Vector Machine), 결정 트리(Decision Tree), 회귀 모델 등</p> <p>-비지도 학습(Unsupervised learning): 입력 값만을 이용하여 학습을 시키는 것 군집화(Clustering), 압축(Compression) 이나 변환 함수를 자동으로 알아내는 등의 문제에 쓰인다</p> <p>.</p>		



Ex) 소셜 네트워크 데이터를 활용하여 사람들 사이의 친함과 친하지 않음의 관계를 예측하는 것

-강화 학습(Reinforcement Learning): 현재의 상태(State)에서 어떤 행동(Action)을 취하는 것인지 학습하는 것으로 행동을 취할 때마다 외부 환경에서 보상을 주는 방식으로 진행된다.

Ex) 체스 게임 프로그램: 현재 나와 적의 말의 배치가 상태가 되고 여기서 어떤 말을 어떻게 움직일지가 행동이 된다.

특징 공간이란 관측 값들이 있는 공간이다. 이것은 여러 차원으로 구성 될 수 있으며 특징 변수가 N개 이면 N차원의 특징 공간을 가진다.

-d 차원 데이터의 특징 벡터를 표현하려면? $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$

질문 내용

수업 시간에 차원과 관련된 설명 중에 “정육면체 안에 구를 넣고, 정육면체의 크기를 무한히 증가시키면 구안에서의 정보들이 없어진다.” 라는 예시를 들었습니다. 이 부분에 대하여 차원의 저주(curse of dimensionality)와 연관시켜 보다 깊이 이해하고 싶습니다.