신경망 지도 학습

2018.07.04

|  |  |  |
| --- | --- | --- |
| 작성자 | 검토자 | 승인자 |
| 이가영 |  |  |

© Copyright 2018 Medicisoft Co.,Ltd.

[1. 지도 학습 2](#_Toc518489369)

[2. 회귀분석(Regression) 3](#_Toc518489370)

[3. 분류(Classification) 4](#_Toc518489371)

[4. 참고문헌 5](#_Toc518489372)

# 지도 학습

##### **정의**

* 특정 입력(input)에 대하여 올바른 정답이 있는 데이터 집합이 주어지는 경우의 학습

##### 학습 방법

* (데이터(data), 레이블(label)) 형태로 학습을 진행한다.

##### 예시 그림 1) 지도학습 예시

* 28 \* 28 크기의 이미지인 MNIST data set이 있으면, 이를 이용해 학습을 진행할 때, training set은 다음과 같이 구성된다.  
  (0을 나타내는 28×28 이미지, 0), (7을 나타내는 28×28 이미지, 7), (6을 나타내는 28×28 이미지, 6), (0을 나타내는 28×28 이미지, 0), …
* 이렇게 구성된 트레이닝 데이터 set으로 학습이 끝나면, 레이블이 지정되지 않은 테스트 데이터셋(test set)을 이용해서, 학습된 알고리즘이 얼마나 정확히 예측(prediction)하는지를 측정할 수 있다.
* 예를 들어 (4를 나타내는 28 \* 28 이미지)를 학습된 분류기에 집어 넣으면, 올바르게 4를 예측 하는지(True Prediction) 아니면 3이나 5와 같은 잘못된 레이블을 예측하는지 (False Prediction) 측정할 수 있다.
* 이 때, 예측하는 결과값이 discrete value(이산값)면 classification(분류)문제   
  - 이 이미지에 해당하는 숫자는 1인가 2인가?
* 예측하는 결과값이 continuous value(연속값)면 regression(회귀)문제   
  - 3개월 뒤 이 아파트 가격은 2억 1천 만원 일 것인가? 2억 2천 만원 일 것인가?

# 회귀분석(Regression)

##### 정의

* Continuous한 연속적인 값을 찾는 것이다.  
  위의 예제에서 보여지는 것과 같이 output은 이산적이고 스칼라 값일 수도 있지만, input에 대응하는 output을 분석하여 연속함수를 찾는 과정을 바로 regression이라고 한다.

##### 예시

그림 2) classification 예시

* 집의 가격에 대해 예측을 해본다고 가정을 한다. 수집한 데이터 집합(Data Set)에 따라 작성한 그래프는 가로 축이 집의 크기이며, 세로축은 집들의 가격이다. 이러한 상황에서 어떤 사람이 750 feet의 집을 판매한다고 가정할 때 얼마를 얻을 수 있는 지에 대해서 알고 싶은 상황이다.
* 이 문제를 해결하기 위해서는 각각의 값들을 선형으로 근사하게 잇는 알고리즘을 떠올릴 수 있다. 데이터집합에 대해 선형 함수를 만들어서 750 feet의 크기라면 대략 150 정도를 얻을 수 있음을 알 수 있다.
* 하지만, 선형으로 만드는 알고리즘은 비효율적이기에 2차 함수로 만드는 알고리즘 등을 더 떠올릴 수 있다. 이를 이용한다면 약 200 정도를 얻을 수 있다.
* 이 예시에서는 각각의 집은 크기(input)에 대한 가격(output)이 주어졌고 Supervised Learning은 이러한 입력들을 기반으로 학습하여 새로운 집에 대한 가격을 알아내는 알고리즘을 개발하고 그 가격을 제시한다.
* 즉, 프로그램은 이러한 정보들을 바탕으로 Input과 Output에 대한 관계를 유추하여 올바른 해답을 제시한다.
* 집의 크기(input)에 해당하는 적절한 집의 가격(output)을 찾는 것이 Regression이다.

# 분류(Classification)

##### 정의

* Input에 대응하는 output을 분석하여 Discrete(이산적인) 값을 찾는 것이다.

##### 예시

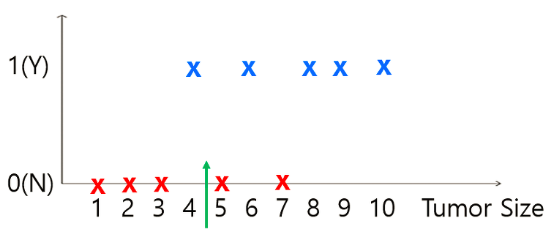


그림 3) classification 예시

* 그림 2는 특정한 종양의 크기에 대하여 그 종양이 악성인지 양성인지를 진단하는 그래프이다. 이 때 4~5 사이의 모호한 종양 크기가 악성인지 양성인지를 구한다고 할 때, 결국 우리가 원하는 결과도 양성 또는 악성의 이산적인 값으로 표현이 된다.

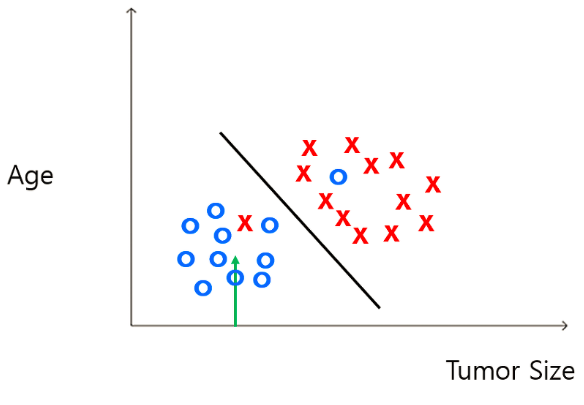


그림 4) classification 예시

* 하지만 단순히 종양의 크기만으로 진단을 내리는 것은 위험하며 같은 크기의 종양이라고 하더라도 나이가 많으면 더 위험할 수 있다. 따라서 종양의 크기와 나이 그리고 악성인지를 나타내는 빨간 색의 X와 양성인지를 나타내는 파랑색의 O의 3가지 속성을 이용한 그래프를 활용하여 판단을 내린다고 가정을 한다. 여기서 학습 알고리즘은 하나의 직선을 그어 양성과 악성을 구분하여 특정 위치의 종양은 양성일 확률이 높음을 판단 할 것이다.

# 참고문헌

##### [Machine Learning] 지도 학습과 비지도 학습 <http://mangkyu.tistory.com/32>

##### 머신러닝(Machine Learning) 알고리즘 분류 – 지도 학습(Supervised Learning), 비지도 학습(Unsupervised Learning), 강화 학습(Reinforcement Learning) <http://solarisailab.com/archives/1785>