# 머신 러닝 보고서

김건호

서울과학고

## 목차



## 1 기초 개념

Regression과 Classification은 유사하면서도 차이가 있다. 우선 둘 모두 Supervised Learning의 방법이기에 어떤 label이 된 독립 변수와 종속 변수가 있는 데이터 셋을 기준으로 새로운 독립 변수를 받았을 때 나오는 종속 변수를 예측하는 것이다. 차이점은 종속 변수의 타입이 되는데, 연속적인 값들을 가질 때를 Regression이라 하고 불연속적이고 이산적인 데이터 형태를 가질때를 Classification이라 한다.

이것 이외에도 Clustering

## 2 Iris Classification

- 한 꽃에 해당하는 값 (2차원 배열에서 가로 행): Instance/Observation
- 한 성질에 해당하는 값 (2차원 배열에서 세로 열): Attribute/Measurement/Dimension
- 예측하려고 하는 값: Target

이 중 나중에 모델을 테스트하기 위해서 Training Data와 Testing Data로 나눈다. Training Data가 많을 수록 모델에게 좋기는 하지만, 가지고 있는 모든 데이터를 Training Data로 사용하면 이 모델이 정확하게 판단할 수 있는지에 대해서 전혀 근거가 없어지기 위해 한다. 따라서 처음 Training Data를 통해서 훈련시키는 과정을 `fit`이라고 하고, Testing Data로 평가하는 과정을 `score`, 또 새로운 데이터를 입력하는 것을 `predict`라고 한다. 이 세 가지 과정 (함수)는 Supervised learning에서 핵심적인 과정이다.

또한 고른 Testing Data가 특별하였을 수도 있기 때문에 모든 가능한 Testing Data / Training Data에 대해서 반복하는 것을 교차 검증이라고 한다.

`sklearn`에는 여러 학습 모델이 있다. 이 중 이번 Iris의 Classification을 위해서는 $k$-NN을 사용한다.

### 2.1 $k$-NN

$k$-NN은 Classification의 가장 단순한 알고리증 중 하나이다. 새로운 데이터를 받았을 때 기존의 Training Data와의 거리를 비교하여 이 새로운 데이터와 거리가 가장 작은 데이터 $k$개를 고르고, $k$개의 데이터의 Target 값들이 동등한 가치로 투표를 한다고 생각하여 새로운 데이터의 Target을 예측한다.

투표라는 개념을 평균 등 연속적인 값으로도 옮길 수 있다면 이를 통해 Regression에 대한 모델도 만들 수 있다.

## 3 기계 학습 (Machine Learning)

### 3.1 기계 학습이란?

기계 학습을 통해서 학습된 기계는 결국은 함수이다. 우리는 이 함수를 우리가 해결하고자 하는 문제에 최적화된 함수를 만들기 위해서 함수의 paramter를 점차 점차 수정해가며 원하는 함수를 만드는 과정이 필요하다. 이 과정을 기계 학습이라고 한다.

### 3.2 기계 학습의 종류

#### 3.2.1 지도 학습

#### 3.2.2 비지도 학습

주로 레이블된 데이터가 많지 않을 때 사용한다.

1. 준지도 학습
2. 전이학습 — 한 특정한 문제에 대해 데이터가 불충분할 때 다른 유사한 문제에서의 데이터를 가져와 학습한다.
3. 제로샷 학습

#### 3.2.3 강화 학습

### 3.3 선형 회귀법

어떤 점들 $(x_n, y_n)$을 근사하는 일차함수 $y = ax + b$로 근사하고 싶다고 하자. 위에서 언급하였듯이 결국은 최적화 문제이기 때문에 최적화 하려고 하는 **오차 함수**를 정의한다. 선형 회귀법에서는 주로 각각의 $(x_n, y_n)$에 대해서 $|ax_n + b - y_n|^2$을 전부 합한 값을 오차 함수라고 한다.

이 오차 함수를 $a, b$를 통해 최소화 한다면 이를 미분을 통해서 구할 수 있다.

### 3.4 다중 회귀법

여러 변수에 대해서 미분을 통해서 0이 되는 계수를 찾거나 Gradient Descent를 통해 구할 수 있다. Gradient Descent는 비슷하게 미분을 하지만 이를 통해서 변수를 조금씩 줄여 여러 번 반복시켜 최적을 찾는 것이다.

## 4 생각

이 내용에 대해 알기 전에는 인공지능이 어떻게 작동하는 지 몰랐는데 강의를 들으며 인공지능이 바탕으로 하는 기본적인 원리는 단순히 최적화 문제를 것을 알게 되었다.

## Appendix

1. Github Repo: `https://github.com/concinnitas43/MachineLearning-Winter2023`