

실습파일 주소 : [https://github.com/jihami/MachineLearning\\_Inflearn](https://github.com/jihami/MachineLearning_Inflearn)

## 01-머신러닝 개념 및 정의

- 인공지능이라는 큰 범주안에 머신러닝이 있고 머신러닝이라는 범주안에 딥러닝  
→ 인공지능>머신러닝>딥러닝
- 머신러닝을 정의한 두인물
  - **Artur Samuel** : 프로그램이 작성되지 않더라도  
**컴퓨터 스스로 학습할 수 있는 능력을 제공할 수 있는 학문**
  - **Tom Mitchell** : 컴퓨터가 어떤 작업을 하는데에 있어서  
**경험으로부터 학습하여 성능에 대한 측정을 향상시키는 학문**
    - ex) 문자 메세지에서 스팸을 거를 때 작업은 문자들이 스팸인지 아닌지 분류하는 것
    - ex) 바둑게임을 할 때 수많은 바둑게임을 하는 것
      - 작업: 바둑게임 하는 것
      - 성능측정: 다음 바둑게임에서 이길 확률
    - ex) 논문을 다운받는 것은 아님 → 성능향상이 되지 않기 때문

## 02-머신러닝 용어 및 라이브러리 정의

- 모델 : 데이터들에 잘 부합하고

우리가 보지 못했던 데이터들을 가지고 잘 예측할 수 있는 것

- 좋은 모델 : 데이터의 패턴들을 잘 학습한 모델

- 특성 : 데이터를 설명하는 입력 변수

ex) 강아지랑 고양이를 분류할 때의 특성은 눈, 귀 모양 등

- 레이블 : 정답, 예측하는 실제 항목

- 추론 : 학습된 모델을 사용하여 새로운 데이터에 적용하여 예측

- 라이브러리

Numpy : 행렬 정렬

Pandas : 인덱스가 있는 1차원 배열, 논리형까지 모든 데이터 타입 가능

Matplotlib : 시각화 할 때 유용

Scikit-learn : 지도 학습과 비지도 학습에서 모델을 만들 때 사용

NLTK : 자연어 처리에서 자주 사용, 토큰화에서도 자주 쓰임

## 03-선형 회귀

- 지도학습의 대표적인 예시
- 지도학습과 비지도 학습의 차이 : 레이블의 존재여부
- 레이블 : 정답여부
- 회귀 : 연속적인 값을 예측할 때 사용, 평균으로 돌아가다
- 러닝에서 일차방정식 표현법 :  $y=ax+b \rightarrow H(x)=Wx+b$
- 비용함수 **MSE** : 가설이 잘 맞는지 아닌지 검사하기 위해 사용
- $\text{cost}(w) = \frac{1}{m} \sum_{i=1}^m (H(x_i) - y)^2$
- 선형 회귀분석 : 회귀식에서 **mes**를 통해 가장 적합한 **w**와 **b**값을 찾는 것
- 최적의 **w**값 찾기 :  $w := w - \alpha \frac{\partial}{\partial w} \text{MSE}$

## 04-다중 선형 회귀

- 최적의 **b**값 찾기 :  $b := b - \alpha \frac{\partial}{\partial b} \text{MSE}$
  - 기본공식 :
    - $H(x)=Wx+b$
    - $H(x_1, x_2)=W_1x_1+W_2x_2 + b$
  - 어머니의 키를 가지고 자식의 키 유추  $\rightarrow 1$
  - 어머니 **with** 아버지의 키를 가지고 자식의 키를 유추  $\rightarrow 2$
  - 전진선택법 : 상수모형으로부터 시작하여 중요하다고 여겨지는 설명변수 부터 차례대로 모델에 추가
  - 후진제거법 : 가장 덜 영향을 준다고 생각하는 변수를 전체에서 1개씩 제거해 나감
- \*변수들이 많다고 다 사용하지 않아도 괜찮은 변수만 사용하기\*

## 05-선형 회귀 모델 구현/colab 설명

- colab : Goggle Colaboratory 약자, 주피터+Google Drive
- GPU : 그래픽 연산을 빠르게 처리하여 결과값을 모니터에 출력하는 연산장치
- TPU : 구글에서 발표한 데이터분석 및 딥러닝용 하드웨어
- 런타임 - 노트설정에서 설정함

## 06-로지스틱 회귀

-로지스틱 회귀 분석 : 반응 변수가 1 또는 0인 이진형 변수에서 쓰이는 회귀 분석 방법

$$-p = ax + b$$

-좌변 : 0 or 1 / 우변 : 마이너스 무한대 ~ 플러스 무한대

- $p$  대신에 Odds 넣기

$$-Odds : \frac{\text{성공 확률}}{\text{실패 확률}} = \frac{\text{성공 확률}}{1 - \text{성공 확률}} = \frac{p}{1-p}$$

$$-Odds에 로그 씌우기 : \log_e(Odds) = \log_2\left(\frac{p}{1-p}\right) = ax + b$$

$$-\text{그래서 결국 식} : p = \frac{e^{ax+b}}{1+e^{ax+b}}$$

## 07-로지스틱 회귀 모델 구현

-실습

## 08-랜덤포레스트