머신러닝

지도학습

→ 지도 학습은 **label이 지정된 데이터 세트(=정답이 있는 데이터)**를 사용하는 머신러닝 접근 방식이다. label이 지정된 데이터 세트를 이용해 데이터를 분류하거나 결과를 정확하게 예측하도록 알고리즘을 설계한다. 인공지능 모델은 label이 있는 입력 및 출력을 사용하여 정확도를 측정하고 학습할 수 있다.

지도 학습은 **회귀(Regression)**와 **분류(Classification)**라는 두 가지 유형으로 나눌 수 있음.

- **회귀(Regression)**

회귀는 알고리즘을 사용하여 종속 변수와 독립 변수 간의 관계를 이해하는 지도 학습의 방법 중 하나이다. 회귀 모델은 특정 비즈니스에 대한 판매 수익 예측과 같이 다양한 데이터 요소를 기반으로 숫자 값을 예측하는 데 유용하다.

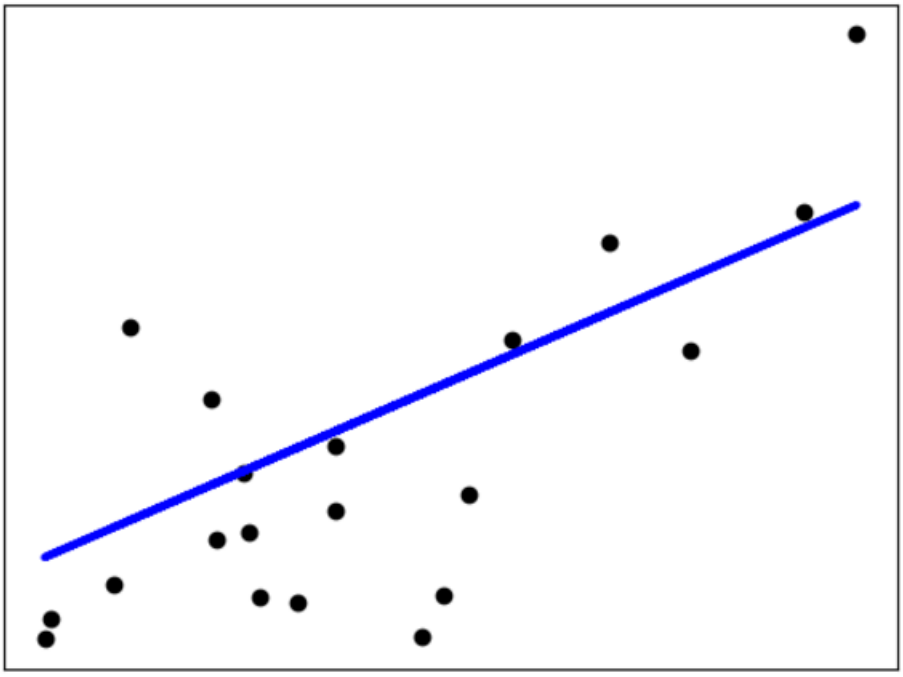
회귀 알고리즘에는 **선형 회귀(Linear regression)**, **로지스틱 회귀(Logistic regression)** 등이 있다.

- 선형 회귀(Linear regression)

머신러닝의 목적은 데이터의 알려진 속성들을 학습하여 예측 모델을 만드는데에 있는데 이때 찾아 낼 수 있는 가장 직관적이고 간단한 모델은 **선(line)!!**

→ 선형회귀란 **데이터를 가장 잘 대변하는 최적의 선을 찾은 과정**

→ 일차 함수($y=ax+b$) 형태로 나타냄. (x = 독립 변수, y = 종속 변수)



- 로지스틱 회귀(Logistic regression)

Logistic → 0 or 1, **Regression →** Fitting이라고 생각하면 편하다.

로지스틱 회귀 모델은 일종의 확률 모델로서 독립 변수의 선형 결합을 이용하여 사건의 발생 가능성을 예측하는데 사용되는 통계기법이며 종속 변수가 범주형 데이터를 대상으로 하며 입력 데이터가 주어졌을 때 해당 데이터의 결과가 특정 분류로 나뉘기 때문에 일종의 분류(classification) 기법이기도 하다.

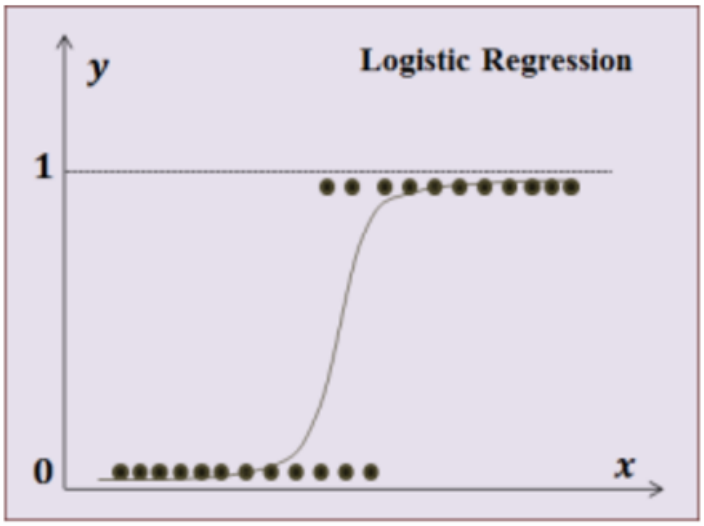
ex) 제품이 불량인지 양품인지 분류, 이메일이 스펨인지 정상메일인지

이러한 판단을 0과 1로 구분, 구한 선형 회귀 식에 독립 변수를 대입하여 나온 값이 0.5(threshold=분류기준)를 넘으면 1, 넘지 않으면 0이라고 판단하는게 일반적인 예시.

$$sigmoid$$

$$y = \frac{1}{1 + e^{-(WX+b)}}$$

나이	질병유무	나이	질병유무	나이	질병유무
22	0	40	0	54	0
23	0	41	1	55	1
24	0	46	0	58	1
27	0	47	0	60	1
28	0	48	0	60	0
30	0	49	1	62	1
30	0	49	0	65	1
32	0	50	1	67	1
33	0	51	0	71	1
35	1	51	1	77	1
38	0	52	0	81	1



위의 시그모이드 함수의 W(weight), b(bias)를 조정하여 이 함수를 점들에 맞게 Fitting 시킴. → 그럼 이 데이터에 맞는 모델 만들어지고, 여기에 다른 값을 넣어도 이 함수를 이용하여 예측값 만들어내는것이 가능해짐.

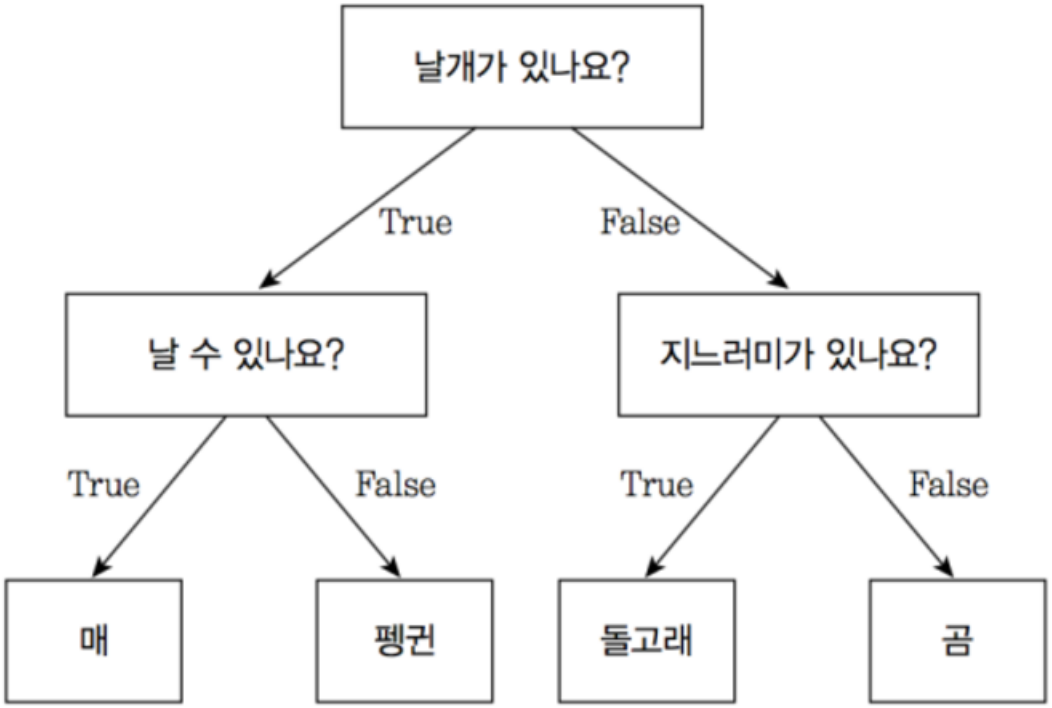
- **분류 (Classification)**

분류 문제는 알고리즘을 사용하여 사과와 오렌지를 분리하는 것과 같이 데이터를 특정 카테고리로 할당하는 방식이다. 또는 지도 학습 알고리즘을 사용하여 받은 편지함과 별도의 폴더에 스팸을 분류할 수 있다.

선형 분류기(Linear classifier), SVM(Support Vector Machine), 의사 결정 트리(Decision tree) 및 랜덤 포레스트(Random forest)등의 분류 알고리즘이 있다.

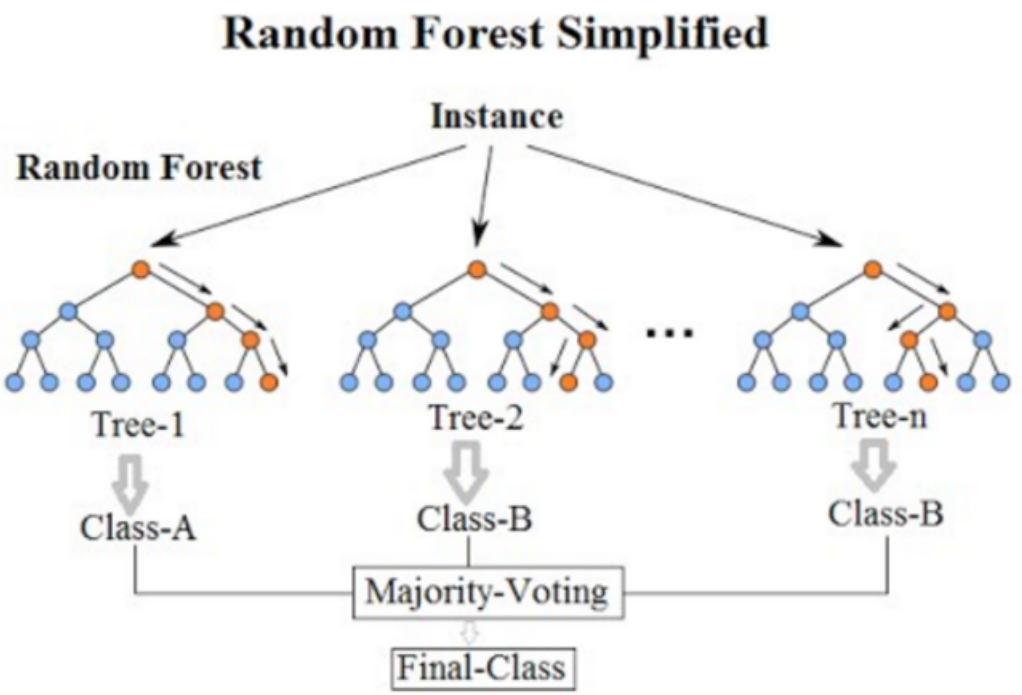
- 의사 결정 트리(Decision tree)

데이터에 있는 규칙을 **학습을 통해 자동으로 찾아내 트리 기반의 분류 규칙을 만드는 것**이다. 일반적으로 가장 쉽게 표현하는 방법은 TRUE/FALSE 기반임.



- 랜덤 포레스트(Random forest)

앙상블(Ensemble) 학습 방법(→ 여러 개의 기본 학습 모델을 조합하여 더 강력한 모델을 만드는 기법) 중 하나로, 여러 개의 의사 결정 트리(Decision tree) 모델을 조합하여 더 강력한 분류 모델을 구축하는 방법이다.



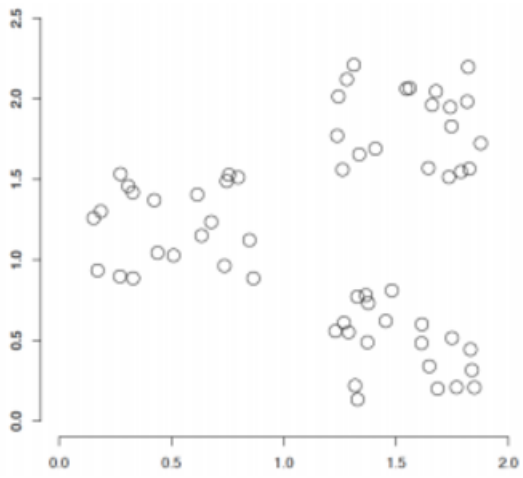
비지도학습

→ 비지도 학습은 머신러닝 알고리즘을 사용하여 **label이 지정되지 않은 데이터 세트(=정답이 없는 데이터)**를 분석하고 클러스터링한다. 이러한 알고리즘은 사람의 개입 없이 데이터에서 숨겨진 패턴을 발견하는 방식이다.

- **클러스터링**

클러스터링(Clustering)이란 샘플 내의 대상들을 일정하게 분류하는 비지도학습.

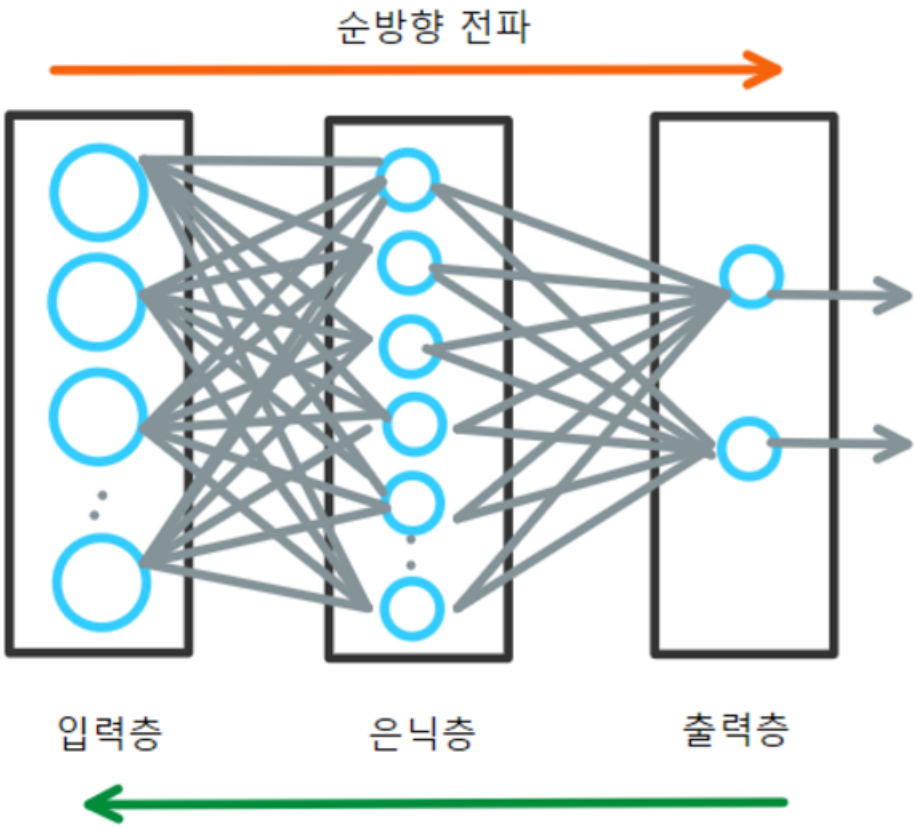
ex) 아래와 같은 2차원 변수 공간에 샘플들이 분포하고 있을 때, 샘플들을 각각의 집단으로 묶어내는 작업이다.



클러스터링 기법 : **K-means Clustering, Hierarchical Clustering, 덴드로그램(Dendrogram) 등등..**

// mlp(살짝)

MLP(Multi Layer Perceptron)란 **여러 개의 퍼셉트론 뉴런**을 **여러 층**으로 쌓은 다층신경망 구조이며,
입력층과 출력층 사이에 하나 이상의 은닉층을 가지고 있는 신경망이다.



초록색 화살표 : 역방향 전파

→ 머신러닝이라기보다는 딥러닝에 가까워서 가볍게 말하고 패스..!