가장 기본적인 계산원리 2가지 선형 회귀와 로지스틱 회귀

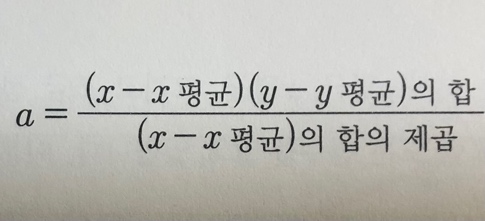
선형 회귀 (Linear Regression)

* 가장 훌륭한 예측선 긋기
* 독립 변수 x 와 종속 변수 y가 있을 때,
* 독립 변수를 사용해 종속 변수의 움직임을 예측하고 설명하는 작업
* 독립 변수 x 가 하나일 때는 단순 선형 회귀
* 독립 변수가 여러 개 있을 때는 다중 선형 회귀라고 한다

기존 데이터(정보)를 가지고 어떤 선이 그려질지를 예측한 뒤, 아직 답이 나오지 않은 무언가를 선에 대입해보는 것

최소 제곱법 (method of least squares)

* 정확한 기울기 a와 정확한 y 절편의 값 b를 알아내는 방법
* 최소 제곱법을 사용하면 a와 b를 바로 구할 수 있다
* 회귀 분석에서 사용되는 표준 방식이다

 텍스트, 화이트보드이(가) 표시된 사진

자동 생성된 설명

이것을 사용하여 오차가 가장 적고 주어진 좌표의 특성을 가장 잘 나타내는 직선을 찾을 수 있다.

위의 내용을 코딩으로 구현하면:

Import numpy as np

X = [2,4,6,8]

Y = [81,93,91,97]

mx = np.mean(X)

my = np.mean(Y)

divisor = sum([(mx – i) \*\* 2 for I in X])

def top(X, mx, Y, my):

d = 0

for I in range(len(X)):

d += (x[i] -mx) \* (y[i] – my)

return d

dividend = top(X,mx,Y,my)

a = dividend / divisor # 기울기 a

b = my – (mx\*a) # y절편 b

하지만 최소 제곱법만으로는 여러개의 입력을 처리하기 힘들어서 사용하기 힘들다.

이렇게 여러 개의 입력 값을 계산할 때에는 임의의 선을 그린 후 이 선이 얼마나 잘 그려졌는지 평가하여 조금씩 수정하는 방법을 사용하는 편이 낫다.

오차 평가 알고리즘이 필요하다. 평균 제곱근 오차(root mean square error)