

# 3장 : 선형회귀

- 선형회귀: 풍부한 예측선 찾기
- 독립변수: 독립적으로 변할 수 있는 값
- 종속변수: 독립변수에 따라 종속적으로 변하는 수
- 단순 선형 회귀: 하나의 변수로 설명 가능
- 다중 선형 회귀: 여러개의 변수 -
- 최소 제곱법

$y = ax + b$  에서  $a$ 와  $b$ 를 구하는 공식 (예측값)

$$a = \frac{(x - x_{\text{평균}})(y - y_{\text{평균}}) \text{의 합}}{(x - x_{\text{평균}})^2 \text{의 합}}$$

(참고)

```

1 import numpy as np
2
3 # x 값과 y값
4 x=[2, 4, 6, 8]
5 y=[81, 93, 91, 97]
6
7 # x와 y의 평균값
8 mx = np.mean(x)
9 my = np.mean(y)
10 print("x의 평균값:", mx)
11 print("y의 평균값:", my)
12
13 # 기울기 공식의 분모
14 divisor = sum([(mx - i)**2 for i in x])
15
16 # 기울기 공식의 분자
17 def top(x, mx, y, my):
18     d = 0
19     for i in range(len(x)):
20         d += (x[i] - mx) * (y[i] - my)
21     return d
22 dividend = top(x, mx, y, my)
23
24 print("분모:", divisor)
25 print("분자:", dividend)
26
27 # 기울기와 y 절편 구하기
28 a = dividend / divisor
29 b = my - (mx*a)
30
31 # 출력으로 확인
32 print("기울기 a =", a)
33 print("y 절편 b =", b)

```

x의 평균값: 5.0  
y의 평균값: 90.5  
분모: 20.0  
분자: 46.0  
기울기 a = 2.3  
y 절편 b = 79.0

## <오차 바로잡기>

-각 점과 그래프 사이의 거리 재기

$$\text{오차} = \text{예측값} - \text{실제 값}$$

↓ 음수 제거

$$\text{오차의 합} = \sum_i (y_i - \hat{y}_i)^2$$

↓

$$\text{평균 제곱 오차 (MSE)} = \frac{1}{n} \sum (y_i - \hat{y}_i)^2$$

## (실습)

```
1 import numpy as np
2
3 #가상의 기울기 a와 y 절편 b
4 fake_a_b=[3,76]
5
6 # x 값과 y값
7 data = [[2, 81], [4, 93], [6, 91], [8, 97]]
8 x = [i[0] for i in data]
9 y = [i[1] for i in data]
10
11 # y=max + b에 a,b 값 대입하여 결과를 출력하는 함수
12 def predict(x):
13     return fake_a_b[0]*x + fake_a_b[1] ) y=actual
14
15 # MSE 함수
16 def mse(y, y_hat):
17     return ((y - y_hat) ** 2).mean() ) 1/n * sum(y_i - y_hat_i)^2
18
19 # MSE 함수를 각 y값에 대입하여 최종 값을 구하는 함수
20 def mse_val(y, predict_result):
21     return mse(np.array(y), np.array(predict_result))
22
23 # 예측값이 들어갈 빈 리스트
24 predict_result = []
25
26 # 모든 x값을 한 번씩 대입하여 predict_result 리스트완성.
27 for i in range(len(x)):
28     predict_result.append(predict(x[i]))
29     print("공부시간=%f, 실제점수=%f, 예측점수=%f" % (x[i], y[i], predict(x[i])))
30
31 # 최종 MSE 출력
32 print("MSE 최종값: " + str(mse_val(predict_result,y)))
```

공부시간=2, 실제점수=81, 예측점수=82  
공부시간=4, 실제점수=93, 예측점수=88  
공부시간=6, 실제점수=91, 예측점수=94  
공부시간=8, 실제점수=97, 예측점수=100  
MSE 최종값: 11.0 → 최종 오차값