Part. 02

회귀분석

| 회귀계수추정

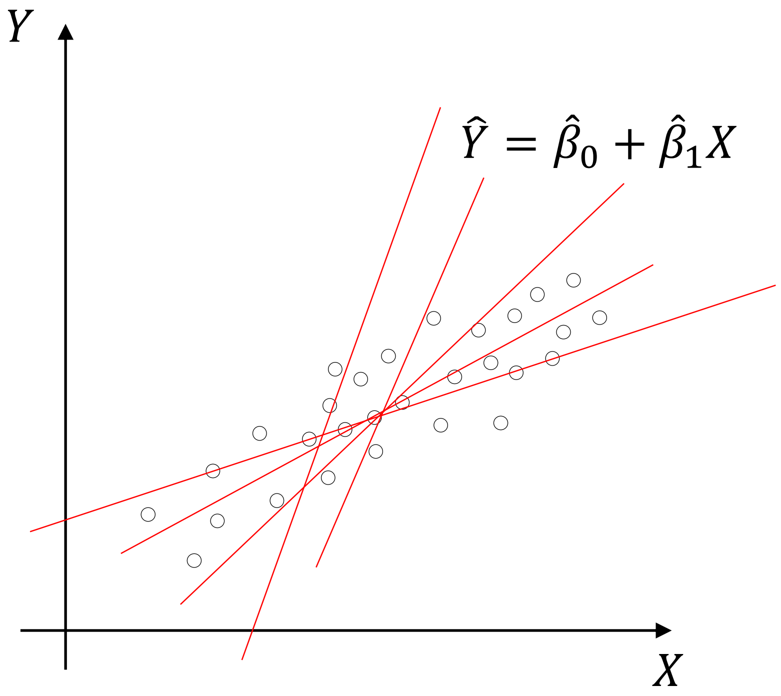
FASTCAMPUS
ONLINE
머신러닝과 데이터분석 A-Z|

강사. 이경택

I 회귀계수추정

- 단순 선형 회귀분석
  - 어떻게 추정 할까?

    - 여러 개의 직선 중 가장 좋은 직선은?



➢ <u>직선과 데이터의 차이가 평균적으로 가장 작아지는 직선</u>

I 회귀계수추정

- 어떻게 추정 할까?

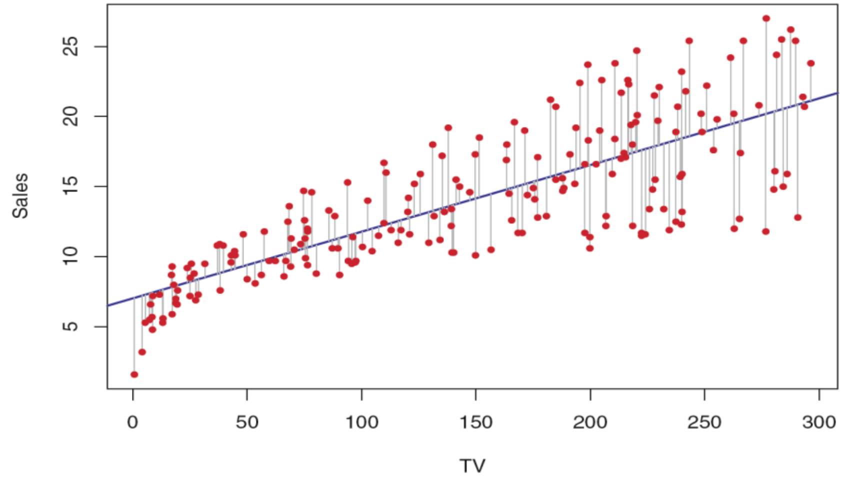
실제 값과 우리가 추정한 값의 차이가 적으면 적을 수록 좋을 것

$$e_i = y_i - \hat{y}_i$$

실제 값과 우리가 추정한 값의 차이를 잔차(residual)라고 하며 이를 최소화 하는 방향으로 추정

- 잔차를 그림으로 나타내면 오른쪽 그림과 같음
- 잔차의 제곱합(SSE; Error Sum of Squares)는 아래와 같이 표현 가능

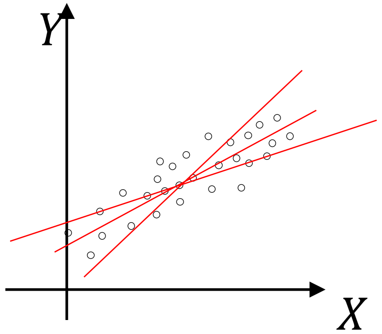
$$SSE = \sum_{i=1}^{n} e_i^2 = e_1^2 + e_2^2 + \cdots + e_n^2$$



I 회귀계수추정

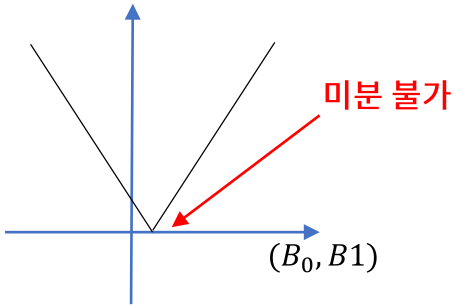
- 굳이 잔차의 제곱합을 최소화 시키는 이유
- 잔차의 합이 0이 되는 해는 무수히 많음 (유일한 해를 찾지 못함)

$$\sum_{i=1}^{n} e_i = e_1 + e_2 + \cdots + e_n = 0$$



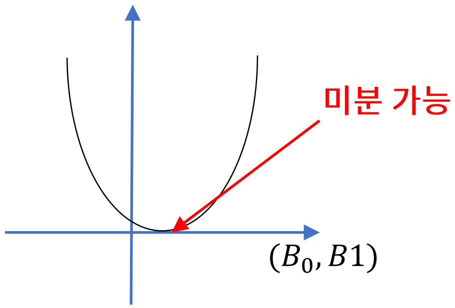
- 잔차의 절대값의 합은 미분이 불가능한 형태

$$\sum_{i=1}^{n} |e_i| = |e_1| + |e_2| + \cdots + |e_n|$$



- 잔차의 제곱 합은 미분이 가능한 형태로 유일한 해를 찾을 수 있음

$$SSE = \sum_{i=1}^{n} e_i^2 = e_1^2 + e_2^2 + \cdots + e_n^2$$

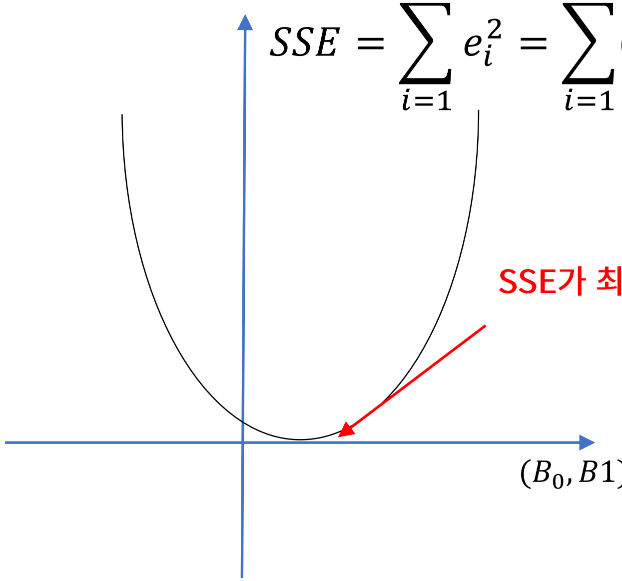


I 회귀계수추정

회귀 계수의 추정

- SSE $\hat{\beta}_0$과 $\hat{\beta}_1$로 편미분하여 연립방정식을 푸는 방법(Least Square Method)

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$



SSE가 최소가 되는 지점은 기울기가 0이 되는 지점

$(B_0, B1)$

$$\frac{\partial L}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial L}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i = 0$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

FAST CAMPUS ONLINE

이경택 강사.

I 회귀계수추정

회귀 계수의 추정

$$\frac{\partial L}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial L}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i = 0$$

$$\sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i - n\beta_0 = 0 \longrightarrow \beta_0 = \frac{1}{n}\sum_{i=1}^{n} y_i - \beta_1 \frac{1}{n}\sum_{i=1}^{n} x_i \longrightarrow \therefore \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} {x_i}^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} y_i - \beta_1 \frac{1}{n}\sum_{i=1}^{n} x_i\right) \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} {x_i}^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i + \beta_1 \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 - \beta_1 \sum_{i=1}^{n} {x_i}^2 = 0$$

$$\Rightarrow \beta_1 \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 - \beta_1 \sum_{i=1}^{n} {x_i}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i$$

$$\therefore \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Part. 02

회귀분석

| 회귀계수의 의미

FASTCAMPUS
ONLINE
머신러닝과 데이터분석 A-Z

강사. 이경택