Ordinary Least Squares (OLS) 과정

1. 모델 및 Notation

다중선형회귀모델:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

잔차의 등분산성, 잔차의 정규성, 선형성, 잔차의 독립성 만족 가정 하에, 독립변수로 예측 변수 추정값을 구하는 선형 모델 실제값 y 가 $\mathbf{X}\beta$ 의 독립변수 * 회귀계수 값에 정규성을 따르는 오차 ϵ으로 구성된다 가정. 모델은 독립변수 y 의 추정치 $\hat{y}$ 을 $\mathbf{X}\beta$ 로 계산하는 모델로, 이때 예측을 정확히 하기 위한 회귀계수 추정, 즉 $\boldsymbol{\beta}$ 값을 최적화 방식으로 잔차 제곱이 최소화가 되는 회귀계수를 구하는 OLS 최적화 방식을 사용. 역행렬 계산으로 데이터가 커지면 계산량이 많아지지만, closed form 해를 구한다는 장점이 있다.

Notation:

- **Design Matrix $\mathbf{X}$**:

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{Np} \end{bmatrix}$$

  디자인행렬은 $N \times p$ 행렬로, N 은 데이터 수 (행 수) and p 는 독립변수 수 (열 수).

- **종속변수 $\mathbf{y}$**:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

  $N \times 1$ 크기의 종속변수 벡터. 예측하고자 하는 변수의 데이터가 있는 열 벡터

- **회귀계수 $\boldsymbol{\beta}$**:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

  $p \times 1$ 크기의 회귀계수 벡터. constant 를 포함 시 β_1이 constant 로 취급

- **오차 $\boldsymbol{\varepsilon}$**:

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

  $N \times 1$크기의 오차 벡터.

2. 학습 최적화 목적함수

예측값과 실제값의 차이인 잔차의 제곱 최소화가 목적함수. 잔차벡터:

$$\mathbf{r} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}.$$

$\mathbf{X}\boldsymbol{\beta}$ 가 예측값, y가 실제값 벡터로 둘의 차를 $\mathbf{r}$ 잔차 벡터로 표현한다.

잔차 제곱합 SSE:

$$\|\mathbf{r}\|^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

3. 목적함수 전개

전개:

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}^\top\mathbf{y} - \mathbf{y}^\top\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{y} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}.$$

$\mathbf{y}^\top\mathbf{X}\boldsymbol{\beta}$ 가 스칼라이므로,

$$\mathbf{y}^\top\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{y}.$$

최종 목적함수:

$$J(\boldsymbol{\beta}) = \Big(\mathbf{y}^\top\mathbf{y} - 2\,\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{y} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}\Big).$$

4. $\boldsymbol{\beta}$에 대해 편미분

목적함수(비용함수) $J(\boldsymbol{\beta})$를 $\boldsymbol{\beta}$ 에 대해 편미분 (2 계수 생략):

$$\frac{\partial J(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\mathbf{X}^\top\mathbf{y} \;+\; \mathbf{X}^\top\mathbf{X}\,\boldsymbol{\beta}.$$

5. OLS 최적해 찾기

gradient가 0이 되도록 두기:

$$-\mathbf{X}^\top\mathbf{y} + \mathbf{X}^\top\mathbf{X}\,\boldsymbol{\beta} = \mathbf{0} \quad\Longrightarrow\quad \mathbf{X}^\top\mathbf{X}\,\boldsymbol{\beta} = \mathbf{X}^\top\mathbf{y}.$$

OLS 해 :

$$\boxed{\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\,\mathbf{X}^\top\mathbf{y}.}$$