Linear Regression

ML Instruction Team, Fall 2022

CE Department
Sharif University of Technology

Formulation

- We generally formulate the linear regression model in matrix form:

$$Y = X\beta + \epsilon$$

where the target value y_i can be evaluated by

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_n x_{in} + \epsilon_i$$

- Y represents a vector of length n containing the observed values $Y = (y_1, \ldots, y_m)^T$.
- β is a vector for the parameters $\beta = (\beta_0, \ldots, \beta_n)^T$.

- ϵ is a vector for the errors $\epsilon = (\epsilon_1, \ldots, \epsilon_m)^T$.

- X is a matrix of the features in which the column of ones incorporates the intercept.

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1n} \\ 1 & x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \ldots & x_{mn} \end{pmatrix}$$

Optimization Problem

■ One suitable estimator of β should be the one minimizing the sum of the squared errors $\|\epsilon\|_2^2 = \sum_{i=1}^{m} \epsilon_i^2 = \epsilon^T \epsilon$.

$$\sum_{i=1}^{m} \epsilon_i^2 = \epsilon^T \epsilon = (Y - X\beta)^T (Y - X\beta)$$
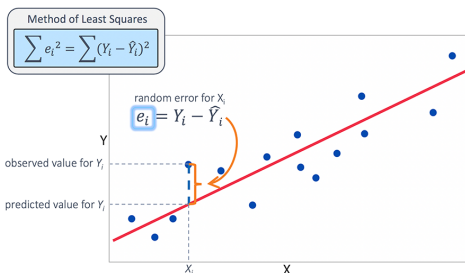
$$= Y^T Y - 2\beta X^T Y + \beta^T X^T X \beta$$

Figure: Simple Linear Regression, Source

Non-Geometric Approach

- Differentiating this term and setting it to zero, we find that the estimate for β, which minimizes the squared error, satisfies the so-called Normal equation:

$$X^T X \hat{\beta} = X^T Y$$

- Provided $X^T X$ is invertible,

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T Y$$

- We can then get the fitted values, $\hat{Y}$, and residuals, $\hat{\epsilon}$,

$$\hat{Y} = X\hat{\beta} = X \left(X^T X\right)^{-1} X^T Y = HY$$

$$\hat{\epsilon} = Y - X\hat{\beta} = Y - \hat{Y} = (I - H)Y$$

where the projection matrix $H = X \left(X^T X\right)^{-1} X^T$.

Geometric Approach

- Another way of looking at this problem is to say we want a solution (our fitted values) that lies in the space spanned by X become as close as possible to Y.

- In this way, the systematic component $X\hat{\beta}$ is the projection of Y onto the space spanned by X, and the residuals are $Y - X\hat{\beta}$.
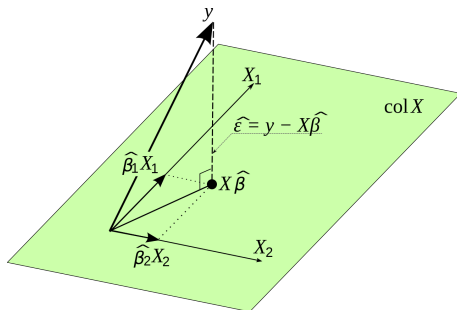

Figure: Geometric Approach of Linear Regression, Source

Extended Versions

■ The Ridge estimate of the linear regression problem is defined as:

$$\hat{\beta}_{\text{Ridge}} = \underset{\beta \in \mathbb{R}^n}{\operatorname{argmin}} \|y - X\beta\|_2^2 + \frac{1}{2}\lambda \sum_{j=1}^{n} |\beta_j|^2$$

$$= \underset{\beta \in \mathbb{R}^n}{\operatorname{argmin}} \|y - X\beta|_2^2 + \frac{1}{2}\lambda \|\beta\|_2^2$$

■ The Lasso estimate of the linear regression problem is defined as:

$$\hat{\beta}_{\text{Lasso}} = \underset{\beta \in \mathbb{R}^n}{\operatorname{argmin}} \|y - X\beta\|_2^2 + \lambda \sum_{j=1}^{n} |\beta_j|$$

$$= \underset{\beta \in \mathbb{R}^n}{\operatorname{argmin}} \|y - X\beta|_2^2 + \lambda \|\beta\|_1$$

Thank You!

Any Question?