

Machine Learning

Haydn Cheng

February 21, 2025

Contents

1	Linear Regression
---	-------------------

2

Linear Regression

In a linear regression model, we assume that the true relationship between the response Y and the predictors $X_j, j = 1, \dots, p$ is given by

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon, \quad (1.1)$$

where ϵ is a random error term which is independent of X_j and has mean zero.¹

From the set of n observation pairs (x_i, y_i) , we estimate the regression coefficients $\hat{\beta}_j, j = 1, \dots, p$ by minimizing the sum of squared residuals (RSS)

$$\text{RSS} = \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad \hat{\mathbf{y}} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p. \quad (1.2)$$

¹Ideally this is truly random but in reality this may be due to predictors that have not been taken into account, which would violate the mean equals to zero property.