Sharif University of Technology
Electrical Engineering Department

Machine Learning and Vision Lab
Pre-Report 2

Amir Hossein Yari
99102507

October 16, 2023

Contents

RSS Matrix Equation Proof

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \quad , \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Np} \end{bmatrix} \quad , \quad \beta = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}$$

$$X\beta = \begin{bmatrix} b_0 + \sum_{j=1}^{p} b_j x_{1j} \\ b_0 + \sum_{j=1}^{p} b_j x_{2j} \\ \vdots \\ b_0 + \sum_{j=1}^{p} b_j x_{Nj} \end{bmatrix} \Rightarrow (y - X\beta) = \begin{bmatrix} y_1 - (b_0 + \sum_{j=1}^{p} b_j x_{1j}) \\ y_2 - (b_0 + \sum_{j=1}^{p} b_j x_{2j}) \\ \vdots \\ y_N - (b_0 + \sum_{j=1}^{p} b_j x_{Nj}) \end{bmatrix}$$

$$(y - X\beta)^T (y - X\beta) = \|y - X\beta\|^2 = \sum_{i=1}^{N} (y_i - (b_0 + \sum_{j=1}^{p} b_j x_{ij}))^2 = \sum_{i=1}^{N} (y_i - f(x_i))^2$$

Correct Equations

In optimal point (β_0^* , β_1^*) the derivative of RSS respect to β_0^* and β_1^* is zero.

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

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

The third equation is correct because:

$$\sum_{i=1}^{n} (y_i - \beta_0^* - \beta_1^* x_i)(x_i - \bar{x}) = \sum_{i=1}^{n} (y_i - \beta_0^* - \beta_1^* x_i) x_i - \bar{x} \sum_{i=1}^{n} (y_i - \beta_0^* - \beta_1^* x_i) = 0 - \bar{x} \times 0 = 0$$

Also the forth equation is correct:

$$\sum_{i=1}^{n} (y_i - \beta_0^* - \beta_1^* x_i)(\beta_0^* + \beta_1^* x_i) = \beta_0^* \sum_{i=1}^{n} (y_i - \beta_0^* - \beta_1^* x_i) + \beta_1^* \sum_{i=1}^{n} x_i (y_i - \beta_0^* - \beta_1^* x_i)$$

$$= \beta_0^* \times 0 + \beta_1^* \times 0 = 0$$

Underfitting

Underfitting occurs when a model is too simple to capture the underlying patterns in the data. It is the opposite of overfitting, which happens when a model is overly complex and fits the training data too closely, but fails to generalize to new, unseen data. Underfit models typically have poor performance on both the training data and new, unseen data. Their predictions are inaccurate because they don't fit the data well.