Shengjie Li
netID: sl1560

CS 536 : Regression and Error

Consider regression in one dimension, with a data set $(x_i, y_i)_{i=1,\cdots,m}$.

1. Find a linear model that minimizes the training error, i.e., $\hat{w}$ and $\hat{b}$ to minimize

$$\sum_{i=1}^{m}(\hat{w}x_i + \hat{b} - y_i)^2.$$

First, let $\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i, \bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i$.

Calculating the partial derivatives and making them equal to zero, we get:

$$2\sum_{i=1}^{m} x_i(\hat{w}x_i + \hat{b} - y_i) = 2\sum_{i=1}^{m}(\hat{w}x_i^2 + \hat{b}x_i - y_ix_i) = 0 \tag{1}$$

$$2\sum_{i=1}^{m}(\hat{w}x_i + \hat{b} - y_i) = 0 \tag{2}$$

Solving (2), we can get:

$$(2) = \hat{w}m\bar{x} + m\hat{b} - m\bar{y} = 0$$
$$\hat{b} = \bar{y} - \hat{w}\bar{x}$$

Put this into (1), we can get:

$$(1) = \hat{w}\sum_{i=1}^{m} x_i^2 + \bar{y}\sum_{i=1}^{m} x_i - \hat{w}\bar{x}\sum_{i=1}^{m} x_i - \sum_{i=1}^{m} y_ix_i = 0$$
$$\hat{w} = \frac{\sum_{i=1}^{m} y_ix_i - \bar{y}\sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} x_i^2 - \bar{x}\sum_{i=1}^{m} x_i} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m}(x_i - \bar{x})^2} = \frac{\mathrm{Cov}(x,y)}{\mathrm{Var}(x)}$$

Thus, $\hat{w} = \frac{\mathrm{Cov}(x,y)}{\mathrm{Var}(x)}, \hat{b} = \bar{y} - \hat{w}\bar{x}$