On Linear Regression

Harry Han

March 2, 2023

1 Notation

1. $\boldsymbol{x}$ is used to denote a vector (or data set)
2. $\bar{\boldsymbol{x}}$ denotes the mean of the vector (data set).

2 The Simplest: $y = ax + b$

Definitio 2.1 (R^2)**.**

For a data set $\boldsymbol{x} = [x_0, x_1, x_2, \cdots, x_{n-1}]^T$ and $\boldsymbol{y} = [y_0, y_1, y_2, \cdots, y_{n-1}]^T$ of the same dimension, there exists a linear regression $y = ax + b$ such that the R^2 value is minimized.

The promised regression is:

$$a = \frac{\boldsymbol{x} \cdot \boldsymbol{y} - n\bar{\boldsymbol{x}}\bar{\boldsymbol{y}}}{(\boldsymbol{x} \cdot \boldsymbol{x}) - n\bar{\boldsymbol{x}}^2};\ b = \frac{(\boldsymbol{x} \cdot \boldsymbol{x})\bar{\boldsymbol{y}} - \bar{\boldsymbol{x}}(\boldsymbol{x} \cdot \boldsymbol{y})}{(\boldsymbol{x} \cdot \boldsymbol{x}) - n\bar{\boldsymbol{x}}^2} \tag{1}$$

notice:

$$b = \bar{\boldsymbol{y}} - a\bar{\boldsymbol{x}}$$