

# On Linear Regression

Harry Han

March 31, 2023

## 1 Notation

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

## 2 The Simplest: $y = ax + b$

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

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

The promised regression is:

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

notice:

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