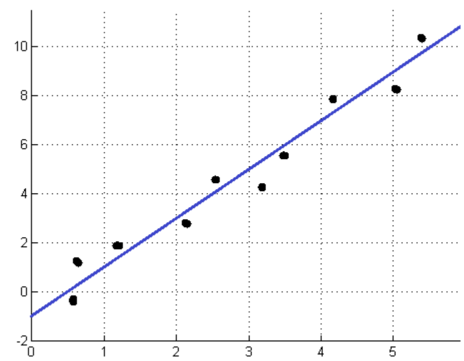
Question 3: Linear Regression
The simplest example of least square is fitting a a straight line to a set of data $\{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\}$.




Let be $y = mx + b$ the equation of the straight line (model), and $R_i = y_i - (mx_i + b)$ for $i = 1, 2, \dots, n$, the error between the model and the observations. We want to minimize

$$R(m, b) = \sum_{i=1}^{n} R_i^2 = \sum_{i=1}^{n} (y_i - mx_i - b)^2 \tag{1}$$

In order to find the minimum of $R$ we calculate the derivative with respect to each variable, we get

$$\frac{\partial R}{\partial m} = -2 \sum_{i=1}^{n} (y_i - mx_i - b)x_i \text{ and } \frac{\partial R}{\partial b} = -2 \sum_{i=1}^{n} (y_i - mx_i - b)$$

Then we need to solve $\frac{\partial R}{\partial m} = 0$ and $\frac{\partial R}{\partial b} = 0$, therefore

$$-2 \sum_{i=1}^{n} (y_i - mx_i - b)x_i = 0 \tag{2}$$

$$-2 \sum_{i=1}^{n} (y_i - mx_i - b) = 0 \tag{3}$$

Notice that equation (2) is equivalent to

$$\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} m(x_i)^2 - \sum_{i=1}^{n} bx_i = 0$$

hence

$$m \sum_{i=1}^{n} (x_i)^2 + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i x_i$$

From equation (3) we get

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} mx_i - \sum_{i=1}^{n} b = 0$$

then

$$m \sum x_i + bn = \sum_{i=1}^{n} y_i$$

Therefore we have the system of equation

$$\begin{aligned} m\sum(x_i)^2 + b\sum x_i &= \sum y_i x_i \\ m\sum x_i + bn &= \sum y_i \end{aligned}$$

By solving for $m$ and $b$ we get

$$m = \frac{n\sum y_i x_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \text{ and } b = \frac{\sum y_i}{n} - m\frac{\sum x_i}{n}.$$

Another approach by using matrix notation. Let be

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, \ X = \begin{bmatrix} m \\ b \end{bmatrix} \text{ and } B = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$$

Minimize (1) is equivalent to

$$\min \|AX - B\|^2 \tag{4}$$

But

$$\|AX - B\|^2 = (AX - B)^T (AX - B)$$

By calculating the derivative of the previous expression with respect to $X$ we get $2A^T(AX - B)$. Therefore in order to solve problem (4) we need to solve

$$A^T(AX - B) = 0 \leftrightarrow A^T AX = A^T B. \tag{5}$$

The system of equation (5) is called normal equations.