Linear Least Squares

We are given a sample of size m of measurements $(z_i, y_i) \in \mathbb{R}^2$ for $i = 1, \ldots, m$. Determine the minimizer c_0 of the problem

$$\min_{c_0 \in \mathbb{R}} \sum_{i=1}^{m} (c_0 - y_i)^2.$$

Hint: This is the simple case where our assumed model is a constant function, i.e., $f(z_i) \equiv c_0$. Set up the design matrix A and solve the normal equation.

Solution:

Model: $f(x) = c_0$, with measurements $(z_i, y_i),\ i = 1, \ldots, m$.

With the design matrix $A = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^{m \times 1} = \mathbb{R}^m$ and $b = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m$, we find

$$\min_{c_0 \in \mathbb{R}} \sum_{i=1}^{m} (c_0 - y_i)^2 = \min_{c_0 \in \mathbb{R}} \|A \cdot c_0 - b\|_2^2.$$

We have

$$A^T A = \sum_{i=1}^{m} 1 = m \quad \text{and} \quad A^T b = \sum_{i=1}^{m} y_i.$$

Thus by solving the normal equation we find

$$A^T A c_0 = A^T b \quad \Leftrightarrow \quad m \cdot c_0 = \sum_{i=1}^{m} y_i \quad \Leftrightarrow \quad c_0 = \frac{1}{m} \sum_{i=1}^{m} y_i.$$

With other words, the best constant fit in the least squares sense is the average of the data.