

Least Squares Fitting of Data

Question: What line $y = c + mx$ best fits the data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$?

In the best of all worlds, all the data points would lie on the line:

$$\begin{aligned}c + mx_1 &= y_1 \\c + mx_2 &= y_2 \\&\vdots \\c + mx_n &= y_n\end{aligned}$$

Considering this as a linear system for unknowns c and m , in matrix form it is

$$\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}}_A \underbrace{\begin{bmatrix} c \\ m \end{bmatrix}}_x = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_b$$