Least Squares Fitting (continued)

There is little doubt that in a system such as this with many equations (think 100 data points, for example) in just two unknowns, the system will be inconsistent. An approximate solution $\hat{c}$ and $\hat{m}$, in the sense of least squares, comes from using projections and the normal equations. In this case

$$A^T A = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

Thus the best fit (in the sense of least squares) is $y = \hat{c} + \hat{m}x$ where

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \hat{c} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$