Least-Square Method

Linear Algebra

Spring 2022

Ramtin Moslemi

Theoretical Explanation

A common task in science and engineering is to analyze and understand relationships among several quantities that vary. Our goal is to find a polynomial relation between two variables x and y that helps us predict future values. We can approximate the value of y, which we'll denote by $\hat{y}$, and it can be computed as a polynomial of degree d:

$$\forall i \in \{1,2,\cdots,n\} \quad | \quad \hat{y}_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_d x_i^d$$

Which could be rewritten in matrix form as:

$$\hat{y} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^d \\ 1 & x_2 & x_2^2 & \cdots & x_2^d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^d \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{bmatrix}$$

We choose notation that is commonly used in the statistical analysis of scientific and engineering data. We refer to X as the **design matrix**, β as the **parameter vector**, and y as the **observation vector**. Statisticians usually introduce a **residual vector** ϵ, defined by $\epsilon = y - \hat{y}$, and write $y = X\beta + \epsilon$.

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\text{observation vector}} = \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^d \\ 1 & x_2 & x_2^2 & \cdots & x_2^d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^d \end{bmatrix}}_{\text{design matrix}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{bmatrix}}_{\text{parameter vector}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{\text{residual vector}}$$

Once X and y are determined, the goal is to minimize the length of ϵ, which amounts to finding a least-squares solution of $X\beta = y$. In each case, the least-squares solution $\hat{\beta}$ is a solution of the normal equations:

$$X^T X\beta = X^T y \rightarrow \hat{\beta} = (X^T X)^{-1} X^T y$$

Hence y could be approximated as such:

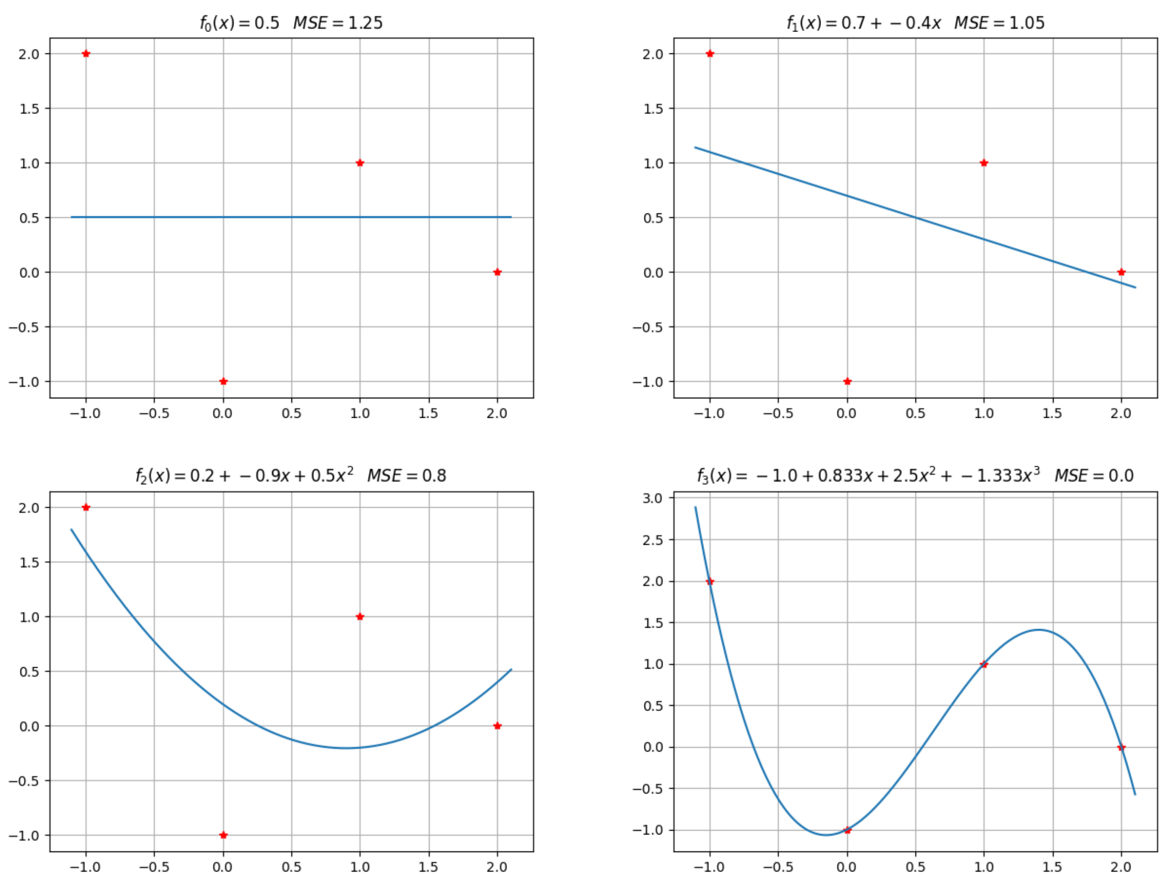
$$\hat{y} = X(X^T X)^{-1} X^T y$$

Practical Explanation

We start by taking the size of the x and y vectors, which we'll store as n, followed by receiving x and y vectors themselves. Now using the last two equations, we will first figure out the parameter vector, otherwise known as the coefficients of the polynomial, and then the mean squared error or MSE for short. We will repeat this for polynomials of all degrees stretching from zero to $n-1$. We stop at $n-1$ since at this point the polynomial will certainly pass through all the points.[1]

After deriving the coefficients, we can finally proceed to drawing the graphs of our polynomials.

An example of four points with coordinates $(-1,2), (0,-1), (1,1), (2,0)$ is shown below:



[1] In linear algebra terms, this happens when the design matrix becomes a square matrix which has the same dimension as observation vector