Least Squares Regression

May 16, 2021

We have to arrays of numbers X and Y. Array X contains independent data points. Array Y contains dependent data points $y_i, i = 1, \ldots, m$.

We want to find $\hat{y}(x)$, that accurately represents given data.

Total squared error is defined as:

$$E = \sum_{i=1}^{m} (\hat{y} - y_i)^2$$

.

The individual errors or residuals are defined as:

$$e_i = (\hat{y} - y_i)$$

.

We try to minimize total squared error and $E = \|e\|_2^2$.

Derivation

Estimation $\hat{y}(x_i)$ for each point x_i:

$$\hat{y}(x_1) = \alpha_1 f_1(x_1) + \alpha_2 f_2(x_1) + \cdots + \alpha_n f_n(x_1),$$
$$\hat{y}(x_2) = \alpha_1 f_1(x_2) + \alpha_2 f_2(x_2) + \cdots + \alpha_n f_n(x_2),$$
$$\cdots$$
$$\hat{y}(x_m) = \alpha_1 f_1(x_m) + \alpha_2 f_2(x_m) + \cdots + \alpha_n f_n(x_m)$$

We can write this system of equations in terms of column vectors $\hat{Y}$ and β:

$\hat{Y}_i = \hat{y}(x_i)$
$\beta_i = \alpha_i$

and mxn matrix A such that it's i-th column equals $F_i(x)$.

The system of equations becomes then: $\hat{Y} = A\beta$

The total squared error is given by E:

$$E = \|\hat{Y} - Y\|_2^2$$

$\hat{Y}$, that is closest to Y is the one that can point perpendicularly to Y .

$$\mathrm{dot}(\hat{Y}, Y - \hat{Y}) = 0$$

$$\hat{Y}^T(Y - \hat{Y}) = 0$$

$$(A\beta)^T(Y - A\beta) = 0$$

$$\beta^T A^T Y - \beta^T A^T A\beta = \beta^T(A^T Y - A^T A\beta) = 0$$

$$A^T Y - A^T A\beta = 0$$

We arrive at the least squares regression formula:

$$\beta = (A^T A)^{-1} A^T Y$$