

Derivation of LS algorithm

Consider a system that can be represented by using a linear-in-parameter function as

$$y = w_0 + w_1 x_1 + \cdots + w_n x_n \quad (1)$$

where w_0 and w_1, \dots, w_n are parameters need to be estimated and x_1, \dots, x_n are input variables.

The basic idea of the LS method is to define $\mathbf{W} = [w_0, w_1, \dots, w_n]^T$, which can satisfy

$$\sum_{i=1}^N [y(i) - \hat{y}(i)]^2 = \min \quad (2)$$

Equation (1) can be written into a matrix form as

$$\mathbf{Y} = \mathbf{X}\mathbf{W} \quad (3)$$

where

$$\mathbf{Y} = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_1(1) & \cdots & x_n(1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_1(N) & \cdots & x_n(N) \end{bmatrix}, \mathbf{W} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}$$

Denote the cost function

$$J(w_0, w_1, \dots, w_n) = (\mathbf{Y} - \mathbf{X}\mathbf{W})^T (\mathbf{Y} - \mathbf{X}\mathbf{W}) = \sum_{j=1}^N \left[y(j) - w_0 - \sum_{i=1}^n w_i x_i(j) \right]^2 \quad (4)$$

The extreme point of (4) exist when

$$\begin{cases} \frac{\partial J(w_0, w_1, \dots, w_n)}{\partial w_0} = -2 \sum_{i=1}^N \left[y(i) - w_0 - \sum_{j=1}^n w_j x_j(i) \right] = -2 \times \mathbf{1}^T (\mathbf{Y} - \mathbf{X}\mathbf{W}) = 0 \\ \frac{\partial J(w_0, w_1, \dots, w_n)}{\partial w_j} = -2 \sum_{i=1}^N x_j(i) \left[y(i) - w_0 - \sum_{j=1}^n w_j x_j(i) \right] = -2 \times \tilde{\mathbf{X}}_j^T (\mathbf{Y} - \mathbf{X}\mathbf{W}) = 0 \end{cases} \quad (5)$$

$$\mathbf{1} = [1, \dots, 1]^T; \tilde{\mathbf{X}}_j = \left[\underbrace{x_j(1), \dots, x_j(N)}_N \right]^T$$

which can be formulated as a matrix form as

$$-\mathbf{X}^T (\mathbf{Y} - \mathbf{X}\mathbf{W}) = -\mathbf{X}^T \mathbf{Y} + \mathbf{X}^T \mathbf{X}\mathbf{W} = \mathbf{0} \quad (6)$$

with $\mathbf{0} = \left[\underbrace{0, \dots, 0}_{n+1} \right]^T$.

Then, the vector \mathbf{W} can be calculated as:

$$\mathbf{W} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (7)$$