Derivation of RLS algorithm

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

$$y = w_0 + w_1 x_1 + \cdots + w_n x_n \tag{1}$$

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

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

$$\sum_{i=1}^{N}\left[y(i) - \hat{y}(i)\right]^2 + \lambda\left(w_0^{\ 2} + \sum_{j=1}^{n} w_j^{\ 2}\right) = \min \tag{2}$$

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

$$\mathbf{Y} = \mathbf{XW} \tag{3}$$

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

$$\begin{aligned} J(w_0, w_1, \cdots, w_n) &= (\mathbf{Y} - \mathbf{XW})^{\mathrm{T}}(\mathbf{Y} - \mathbf{XW}) + \lambda \mathbf{W}^{\mathrm{T}}\mathbf{W} \\ &= \sum_{j=1}^{N}\left[y(j) - w_0 - \sum_{i=1}^{n} w_i x_i(j)\right]^2 + \lambda\left(w_0^{\ 2} + \sum_{j=1}^{n} w_j^{\ 2}\right) \end{aligned} \tag{4}$$

The extreme point of (4) exist when

$$\begin{cases} \dfrac{\partial J(w_0, w_1, \cdots, w_n)}{\partial w_0} = -2\displaystyle\sum_{i=1}^{N}\left[y(i) - w_0 - \sum_{j=1}^{n} w_j x_j(i)\right] + 2\lambda w_0 \\ \quad = -2 \times \mathbf{1}^{\mathrm{T}}(\mathbf{Y} - \mathbf{XW}) + 2\lambda w_0 = 0 \\ \dfrac{\partial J(w_0, w_1, \cdots, w_n)}{\partial w_j} = -2\displaystyle\sum_{i=1}^{N} x_j(i)\left[y(i) - w_0 - \sum_{j=1}^{n} w_j x_j(i)\right] + 2\lambda w_j \\ \quad = -2 \times \tilde{\mathbf{X}}_j^{\ \mathrm{T}}(\mathbf{Y} - \mathbf{XW}) + 2\lambda w_j = 0 \end{cases} \tag{5}$$

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

which can be formulated as a matrix form as

$$-\mathbf{X}^{\mathrm{T}}(\mathbf{Y} - \mathbf{XW}) + \lambda\mathbf{W} = -\mathbf{X}^{\mathrm{T}}\mathbf{Y} + \left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}\right)\mathbf{W} = \mathbf{0} \tag{6}$$

with $\mathbf{0} = [\underbrace{0, \ldots, 0}_{n+1}]^{\mathrm{T}}$.

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

$$\mathbf{W} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} \tag{7}$$