CSE 40625/60625: Machine Learning, fall 2021

Priscila Moreira - pmoreira@nd.edu

Question 1 (2pts). When the linear regression objective function is defined as

$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2$$

we had the solution for $\mathbf{w}$: $\mathbf{w} = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}(\mathbf{\Phi}^T\mathbf{t})$

Adding a regularization term can help to reduce the overfitting issue. The objective function becomes:

$$\widetilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2$$

Then, what's the close form solution for $\mathbf{w}$?
Please show the detailed derivation for it.

The maximization of the likelihood function can be applied to determine $\mathbf{W}$ and such process is equivalent to minimizing a sum-of-squares error function given by $E_D(\mathbf{W})$.

Since the objective function $\tilde{E}(\mathbf{W})$ remains a quadratic function of $\mathbf{W}$ its exact minimizer can be found in closed form.
To do that we can set the gradient of $\tilde{E}(\mathbf{W})$ with respect to $\mathbf{W}$ to zero and solve $\mathbf{W}$.

$$\nabla \tilde{E}(W) = 0$$

$$\nabla\left[\frac{1}{2}\sum_{n=1}^{N}\{y(x_n, w) - t_n\}^2 + \frac{\lambda}{2}\|W\|^2\right] = 0 \qquad y(x_n, w) = \mathbf{W}^T\phi(x_n)$$

$$\nabla\left[\frac{1}{2}\sum_{n=1}^{N}\{\mathbf{W}^T\phi(x_n) - t_n\}^2 + \frac{\lambda}{2}\|W\|^2\right] = 0$$

$$\frac{\cancel{2}}{\cancel{2}}\sum_{n=1}^{N}\{\mathbf{W}^T\phi(x_n) - t_n\}\phi(x_n)^T + \lambda\frac{\cancel{2}}{\cancel{2}}W = 0$$

$$\underbrace{\sum_{n=1}^{N} W^T\phi(x_n)\phi(x_n)^T}_{\Phi^T\Phi W} - \underbrace{\sum_{n=1}^{N} t_n\phi(x_n)^T}_{\Phi^T\mathbf{t}} + \lambda W = 0$$

$$\Phi^T\Phi W - \Phi^T\mathbf{t} + \lambda W = 0$$

$$\Phi^T\Phi W + \lambda W = \Phi^T\mathbf{t}$$

$$(\Phi^T\Phi + \lambda I)W = \Phi^T\mathbf{t}$$

$$\underbrace{\boxed{(\Phi^T\Phi + \lambda I)^{-1}(\Phi^T\Phi + \lambda I)}}_{I} W = (\Phi^T\Phi + \lambda I)^{-1}\Phi^T\mathbf{t}$$

$$\mathbf{W} = (\Phi^T\Phi + \lambda I)^{-1}\Phi^T\mathbf{t}$$ → close form solution for $\mathbf{W}$

Design Matrix

$$\underline{\Phi} = \begin{pmatrix} \phi_0(x_1) & \phi_1(x_1) & \dots & \phi_{M-1}(x_1) \\ \phi_0(x_2) & \phi_1(x_2) & \dots & \phi_{M-1}(x_2) \\ \vdots & \vdots & & \vdots \\ \phi_0(x_N) & \phi_1(x_N) & \cdots & \phi_{M-1}(x_N) \end{pmatrix}_{N\times M}$$

Target vector

$$\underline{\mathbf{t}} = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}_{N\times 1}$$