Kernel Ridge Regression

Saturday, January 20, 2018 3:52 PM

The objective is

$$\min_{w} \frac{1}{2} \sum_{n=1}^{N} \left(w^T \phi(x_n) - t_n \right)^2 + \frac{\lambda}{2} w^T w$$

To solve this problem, we take the derivative with respect to W and set it to 0.

$$\frac{\partial}{\partial w} \left[\frac{1}{2} \sum_{n=1}^{N} \underbrace{\left(w^T \phi(x_n) - t_n \right)^2}_{\text{Let's expand and simplify}} + \frac{\lambda}{2} w^T w \right]$$

scalar

$$(w^T \phi(x_n) - t_n)^T (w^T \phi(x_n) - t_n)$$

$$(\phi(x_n)^T w - t_n^T)(w^T \phi(x_n) - t_n)$$

$$\phi(x_n)^T w w^T \phi(x_n) - t_n^T w^T \phi(x_n) - \phi(x_n)^T w t_n + t_n^T t_n$$

$$\phi(x_n)^T w w^T \phi(x_n) - 2 t_n w^T \phi(x_n) + t_n^2$$

or

$$w^T \phi(x_n) \phi(x_n)^T w - 2 t_n w^T \phi(x_n) + t_n^2$$

With the simplified version, the new formulation.

$$\frac{\partial}{\partial w}\left[\frac{1}{2} \sum_{n=1}^{N} w^T\phi(x_n)\phi(x_n)^T w - 2t_n w^T\phi(x_n) + t_n^2 \quad + \frac{\lambda}{2} w^T w \right]$$

Now we can take the derivative

$$\frac{1}{2}\sum_{n=1}^{N} 2\phi(x_n)\phi(x_n)^T w - 2t_n\phi(x_n) \quad + \lambda I w = 0$$

$$\sum_{n=1}^{N} \phi(x_n)\phi(x_n)^T w - t_n\phi(x_n) + \lambda I w = 0$$

$$\left[\sum_{n=1}^{N} \phi(x_n)\phi(x_n)^T\right] w + [\lambda I] w = \sum_{n=1}^{N} t_n \phi(x_n)$$

Let's define $\Phi = \begin{bmatrix} \phi(x_1)^T \\ \phi(x_2)^T \\ \vdots \end{bmatrix}$ and $t = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \end{bmatrix}$

$$\Phi^T \Phi w + [\lambda I] w = \Phi^T t$$

$$(\Phi^T \Phi + \lambda I) w = \Phi t$$