

Machine Learning 6.867 - Pset 1

September 27, 2015

1 Ridge Regression

1.1 Implementation

Ridge regression is the particular case of regularized least squares with a quadratic regularizer term. The error function that we aim to minimize over is given by:

$$\frac{1}{2} \sum_{n=1}^N (t_n - \mathbf{w}^T \phi(\mathbf{x}_n))^2 + \frac{\lambda}{2} \mathbf{w}^T \mathbf{w} \quad (1)$$

The closed-form solution of this problem is well-known, and can be derived by setting the gradient of (1) equal to zero. The optimal solution for \mathbf{w} is provided by Bishop (2006), page 145:

$$\mathbf{w}_{ridge} = (\lambda \mathbf{I} + \Phi^T \Phi)^{-1} \Phi^T \mathbf{t} \quad (2)$$

We coded this method in MatLab and tested our program using data from Bishop Figure 1.4, varying the parameters of λ and M . For the extreme cases, we observed that if $\lambda \leq 0.0001$, then $\mathbf{w}_{ridge} \approx \mathbf{w}_{OLS}$, and if $\lambda \geq 100$, then $\mathbf{w}_{ridge} \approx \mathbf{0}$.