ECE 517 HW 5.1

Joseph Hilland

November 2023

1 Problem

The criterion given in video 1, slide 5 has been extended to the following ridge regression:

$$\text{minimize } J(w) = E(e^2) + \gamma \|w\|^2$$

$J(w)$ is the function to be minimized, $E(e^2)$ is the mean square error, γ is the regularization parameter and lastly $\|w\|^2$ is the squared norm of the parameter vector w.

Next step is to express the MSE and the norm in terms of the data and parameters. The MSE is expressed as:

$$E(e^2) = \|y - Xw\|^2$$

y is the vector of observed values, X is the matrix of input features, w is the parameter vector.

The function then becomes:

$$J(w) = \|y - Xw\|^2 + \gamma \|w\|^2$$

To minimize the function we will take the derivative with respect to w and then set it to zero. The function then becomes.

$$\frac{\partial J(w)}{\partial w} = -2X^T(y - Xw) + 2\gamma w = 0$$

Next solve for w:

$$X^T(y - Xw) + \gamma w = 0$$

$$X^T y - X^T X w + \gamma w = 0$$

$$X^T X w + \gamma w = X^T y$$

$$(X^T X + \gamma I)w = X^T y$$

$$w = (X^T X + \gamma I)^{-1} X^T y$$

So the optimum value of w for the Ridge Regression is:

$$w = (X^T X + \gamma I)^{-1} X^T y$$