Statistical Machine Learning - HW 4

Homework 5 Problem

For leave-one-out cross validation (or equivalently c-fold cross validation with $c = n$) the cross validation error E_{val} for ridge regression actually has a closed-form solution

$$E_{\text{val}} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\hat{y}_i - y_i}{1 - [H(\gamma)]_{ii}} \right)^2, \tag{1}$$

where $\hat{y}_i$ is the prediction of y_i when the model is learned from all n data points (no data point is left out), the regularization parameter γ and $[H(\gamma)]_{ii}$ is element (i, i) of the matrix

$$H(\gamma) = X(X^\top X + \gamma I)^{-1} X^\top.$$

In this problem, we will let $x_i^\top$ denote the entire ith row of X.

(a)

Let X_{-i} denote the matrix X where row i is removed, and y_{-i} is the column vector y with element i removed. Show that

$$X_{-i}^\top X_{-i} = X^\top X - x_i x_i^\top,$$

$$X_{-i}^\top y_{-i} = X^\top y - x_i y_i,$$

and

$$[H(\gamma)]_{ii} = x_i^\top (X^\top X + \gamma I)^{-1} x_i.$$

(b)

Using the results from (a) and a special case of the matrix inversion lemma

$$(A - vv^\top)^{-1} = A^{-1} + \frac{A^{-1}vv^\top A^{-1}}{1 - v^\top A^{-1} v},$$

show that

$$\hat{\beta}_{-i} = \hat{\beta} + \frac{1}{1 - [H(\gamma)]_{ii}} (X^\top X + \gamma I)^{-1} x_i (\hat{y}_i - y_i).$$

Here $\hat{\beta}_{-i}$ are the parameters learned from all data points except i, and $\hat{\beta}$ the parameters learned from all data.

Hint: Start from the ridge regression expression for $\hat{\beta}_{-i}$ as a function of X_{-i}, y_{-i} and γ.

(c)

Use your result from (b) to derive eq. (1), starting from

$$E_{\text{val}} = \frac{1}{n} \sum_{i=1}^{n} (\hat{\beta}_{-i}^\top x_i - y_i)^2.$$

(d)

Describe (in a few sentences) what eq. (1) can be used for in practice.