Saarland University

The Elements of Stastical Learning

Assignement 4

Due Date: 13.12.2017

Thibault Schowing *Mat. 2571837*
Sarah Mcleod *Mat. 2566398*
December 8, 2017

Problem 1

Prove that for linear and polynomial least squares regression, the LOOCV estimate for the test MSE can be calculated using the following formula:

$$CV_{(n)} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{1 - h_i}\right)^2 \tag{1}$$

Where h_i is the leverage (3.37, ISLR p98)

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i'=1}^{n}(x_i' - \bar{x})^2} \tag{2}$$

We first have this equation, that can take long if n is big because it has to fit every model.

$$MSE_i = (y_i - \hat{y}_i)^2$$

$$CV_{(n)} = \frac{1}{n}\sum_{i=1}^{n} MSE_i$$

So knowing that $\hat{y} = Hy$ and as it is a Leave One Out cross validation, we fit n times the model with one element out. So we have:

$$H = X(X^TX)^{-1}X^T$$
$$H^{-i} = X_{-i}(X_{-i}^TX_{-i})^{-1}X_{-i}^T$$

The hat matrix with all the data and with one out, respectively
Then we have

$$\hat{y}_i = x_i^T[X(X^TX)^{-1}X^T]y$$
$$\hat{y}_{-i} = x_i^T[X_{-i}(X_{-i}^TX_{-i})^{-1}X_{-i}^T]y_{-i}$$

The fitted values at x_i when using all the data points and when leaving one out. We can then to the following:

$$\hat{y}_{-i} = \sum_{i \neq j} H_{ij}y_j + H_{ii}\hat{y}_{-i}$$

$$\hat{y}_{-i} = \sum_{j}^{m} H_{ij}y_j - H_{ii}y_i + H_{ii}\hat{y}_{-i}$$

$$\hat{y}_{-i} = \hat{y}_i - H_{ii}y_i + H_{ii}\hat{y}_{-i}$$

We substitute the in the prediction error:

$$y_i - \hat{y}_{-i} = y_i - (\hat{y}_i - H_{ii}y_i + H_{ii}\hat{y}_{-i})$$

$$y_i - H_{ii}y_i - \hat{y}_{-i} - H_{ii}\hat{y}_{-i} = y_i - \hat{y}_i$$

$$y_i - \hat{y}_{-i} = \frac{y_i - \hat{y}_i}{1 - H_{ii}}$$

Taking the mean square error leads to equation 1.

Problem 2

1. Ridge regression is done by minimizing the RSS with a quadratic penalty term:

$$minimize(y - X\beta)^T(y - X\beta) + \lambda\beta^T\beta$$

to show that the solutions take on the form:

$$\hat{\beta}^{ridge} = (X^TX + \lambda yI)^{-1}X^Ty$$

First we expand the equation to:

$$y^Ty - y^TX\beta - yX^T\beta^T + X^T\beta^TX\beta + \lambda\beta^T\beta$$

which simplifies to:

$$y^Ty - yX^T\beta^T - yX^T\beta^T + X^T\beta^TX\beta + \lambda\beta^T\beta$$

and:

$$y^Ty - 2yX^T\beta^T + X^T\beta^TX\beta + \lambda\beta^T\beta$$

We take the first derivative with respect to β, which gives us:

$$0 + -2X^Ty + 2(XX^T)\beta + 2\lambda\beta$$

setting this to zero this can be simplified to:

$$2(XX^T)\beta + 2\lambda\beta = 2X^Ty$$

and further simplified to:

$$((XX^T) + \lambda I)\beta = X^Ty$$

where I is the p x p identity matrix (added to the matrix math works out correctly). Solving for β gives:

$$\hat{\beta}^{ridge} = (X^TX + \lambda yI)^{-1}X^Ty$$

2.

Problem 3

Problem 4