Saarland University

The Elements of Stastical Learning

Assignement 4

Due Date: 13.12.2017

Thibault Schowing *Mat. 2571837*
Sarah Mcleod *Mat. 2566398*
November 30, 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.

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

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

The equation 1 is like the ordinary MSE, except the ith residual is divided by $1 - h_i$. The leverage lies between $1/n$ and 1, and reflects the amount that an observation influences its own fit. Hence the residuals for high-leverage points are inflated in this formula by exactly the right amount for this equality to hold.(ISLR, p.180)

h_{ii} correspond to the i^{th} diagonal element of the hat matrix.

$$h_{ii} = [H]_{ii}$$

$$H = X(X^T X)^{-1} X^T$$

We can now

Problem 2

Problem 3

Problem 4