# Exercises, STK4040, week 45

November 5, 2007

## Exercise 1: MSE

Let $f$ be a regression trained (estimated) on a data set X and $\boldsymbol{y}$, and let $\hat{y} = f(\boldsymbol{x})$ be the predicted response value for a given new observation $\boldsymbol{x}$. Show that

$$\text{MSE}(\hat{y}) = \text{Bias}^2(\hat{y}) + \text{Var}(\hat{y} - Y), \tag{1}$$

where $Y$ is (the unknown) response value corresponding to $\boldsymbol{x}$. (We treat X and $\boldsymbol{x}$ as given.) Note that this is true for regressions in general; we have not assumed anything about the shape of $f$ (for instance that is should be a linear regression).

## Exercise 2: leverage

Assume the model

$$\boldsymbol{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

where X is known and $\boldsymbol{\varepsilon} \sim N_n(0, \sigma^2 \mathrm{I})$.

Given a data set X and $\boldsymbol{y}$. Let $e_i = y_i - \hat{y}_i$ be the residuals from a linear regression of $\boldsymbol{y}$ onto X. Let $h_i$ be the diagonal entries of the 'hat matrix' $\mathrm{H} = \mathrm{X}(\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T$. $h_i$ is called the 'leverage' of $\boldsymbol{x}_i$, and is a measure of how much $\boldsymbol{x}_i$ potentially can affect the regression. Note that $h_i = \boldsymbol{x}_i^T(\mathrm{X}^T\mathrm{X})^{-1}\boldsymbol{x}_i$.

### a)

Let $\mathrm{X}_{(i)}$ and $\boldsymbol{y}_{(i)}$ be X and $\boldsymbol{y}$ with the $i$th observation removed, and let $\boldsymbol{b}_{(i)}$ be the estimate of $\boldsymbol{\beta}$ using $\mathrm{X}_{(i)}$ and $\boldsymbol{y}_{(i)}$. Show that

$$e_{(i)} = y_i - \boldsymbol{x}_i^T \boldsymbol{b}_{(i)} = e_i/(1 - h_i). \tag{3}$$

(In other words: one can calculate the residuals from a full cross-validation (A.K.A. leave one out cross-validation) of a linear regression withoug having to refit the regression for each observation.)

Tip: to save a bit of calculation, you can use the fact that

$$(\mathrm{X}_{(i)}^T \mathrm{X}_{(i)})^{-1} = (\mathrm{X}^T \mathrm{X})^{-1} + \frac{(\mathrm{X}^T \mathrm{X})^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^T (\mathrm{X}^T \mathrm{X})^{-1}}{1 - h_i}. \tag{4}$$

(Extra exercise: Use the fact that $\mathrm{X}_{(i)}^T \mathrm{X}_{(i)} = \mathrm{X}^T \mathrm{X} - \boldsymbol{x}_i \boldsymbol{x}_i^T$ to show (4).)

b)

Assume that $\mathrm{X} = (\mathbf{1}\ \mathrm{X}_1)$ in (2). Let $\mathrm{X}_c$ and $\boldsymbol{y}_c$ be $\mathrm{X}_1$ and $\boldsymbol{y}$ centered, and let $\boldsymbol{y}_c = \mathrm{X}_c \boldsymbol{\beta}_c + \boldsymbol{\varepsilon}$ be the corresponding centered model. Call the leverages for this model $g_i$, and show that

$$h_i = g_i + 1/n, \tag{5}$$

where $n$ is the number of observations, and $h_i$ are the leverages for the observations in model (2). (Tip: Note that if we write the observations in model (2) as $\begin{pmatrix} 1 & \boldsymbol{x}_i \end{pmatrix}$, then the corresponding observations in the centered model can be written as $\boldsymbol{x}_i - \bar{\boldsymbol{x}}$, where $\bar{\boldsymbol{x}}$ is the coloumn means of $\mathrm{X}_1$.)