Elements of Machine Learning
Assignment 3

William LaCroix (7038732)
wila00001@stud.uni-saarland.de

Problem 1 (T, 14 Points). **Cross-Validation.**

1. [4pts] Explain the impact of the value for k in k-fold cross validation. Where does k-fold CV fit in between the validation set approach and LOOCV and what is the advantage of using it?

The k in k-folds CV represents the number of equal partitions, and thus the number of tests to compare against each other, where $\frac{n}{k}$ is the size of the test set. When $k = n$, then it is just LOOCV. When $k = 1$, then the entire dataset is the training set. One common split for validation sets is 80% training set, 20% validation/test set, which corresponds to k=5-fold CV. One major advantage k-fold has over LOOCV, is the amount of computational savings, running the test k times, as opposed to training a model n times. Another benefit is being able to tailor the bias-variance split, where small k values give lower bias, with reducing variance as k approaches n.

2. [2pts] For the hat matrix H, defined as

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

the diagonal element $h_i = H_{(ii)}$ is called the leverage. What is the meaning of the leverage h_i for sample i? Consequently, what effect on model estimation does removing a sample with high leverage from the dataset have?

The leverage represents the expected value's distance from the mean, weighted by the relative local data density. A data point in a relatively sparse section of the plot has higher leverage than those closer to the main body of the data. This is represented by the diagonal element in the hat matrix, and essentially measures the weighted distance from the mean. Removing a point with high leverage shifts a model closer to the rest of the points, establishing a new mean, and hopefully better capturing the underlying distribution.

3. [8pts] Prove that for linear and polynomial least squares regression, the LOOCV estimate for the test MSE can be calculated as

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

Hint: First, properly understand and define all variables that are relevant for the situation where you set sample i aside.

Expected test MSE: $E\left(y_0 - \hat{f}\left(x_0\right)\right)^2$
ISLR 3.37: $h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\Sigma_{i'=1}^{n}(x_{i'} - \bar{x})^2}$
LOOCV 1 test $MSE_i = (y_i - \hat{y}_i)^2$
ISLR 5.1 $CV_{(n)} = \frac{1}{n}\Sigma_{i=1}^{n} MSE_i$
-> ???