Elements of Machine Learning
Exercise Sheet 3
Winter Term 2023/2024

William LaCroix - wila00001@stud.uni-saarland.de - 7038732
Philipp Hawlitschek - phha00002@stud.unisaarland.de - 7043167

Problem 1 (T, 4 Points). **Cross-Validation.**

1. [2pts] 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. Common choices for number of folds are $k \in 5, 10$, and this choice depends largely on dataset size (ESL sec 7.10.1, ISLP p209). 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] Explain how an outlier in a dataset can affect scores of LOOCV. In this setting, can k-fold cross-validation address the drawbacks of LOOCV?

Outliers in a dataset have a large impact in LOOCV, since the model is repeatedly trained on nearly identical training sets. This leads to poor prediction of the left-out data point. Additionally, outliers will not be predicted well by the trained function, so will contribute a large error. LOOCV is approximately unbiased, but suffers from high variance. (ISLP 209)

Conversely, outliers in test data for k-fold CV are tempered somewhat by more regular data, while outliers in training data contribute less to the overall model performance. K-fold CV introduces some bias by training on less of the data set, but this results in significantly lower variance, and broader generalizability. (ISLP 209)