-Following the lecture and textbooks, please answer the following questions An Introduction to Statistical Learning (with python application) Chapter 5 Exercise: (3), (4), (5), (6), (7).

5.3) 3. We now review k-fold cross-validation.
(a) Explain how k-fold cross-validation is implemented.
Answer:
The k-fold cross-validation technique serves as a means for approximating the test error associated with a specific statistical estimator. To implement this method, the initial step involves dividing the training set, which consists of N data points, into k mutually exclusive and approximately equal subsets. In cases where N is not evenly divisible by k, we distribute the data into k folds as evenly as possible, such as when N = 1003 and k = 10, resulting in 3 folds with 101 elements and 7 folds with 100 elements.

Subsequently, the model is fitted k times, with each iteration excluding one of the folds as a validation set while using the remaining k-1 folds for training. During each of these k model fits, the validation error is calculated using the excluded fold. Ultimately, the test error estimate is derived by averaging the validation errors from all k iterations.

(b) What are the advantages and disadvantages of k-fold cross validation
relative to:

i. The validation set approach?

Answer: In comparison to k-fold cross-validation, the validation set approach offers advantages in terms of computational efficiency, as it involves fitting the model only once. It is also a simpler and more straightforward technique to implement. However, it tends to overstate the test error because it utilizes only half of the sample for model fitting. Generally, a larger sample size results in lower test error. Furthermore, relying on only half of the sample introduces dependence on the specific half chosen for estimating the test error.

Similar considerations apply to k-fold cross-validation, although the impact is less pronounced. This difference becomes apparent when observing the results presented in Figures 5.2 and 5.4 in the referenced text, particularly in the right-hand panels.

From book figures:

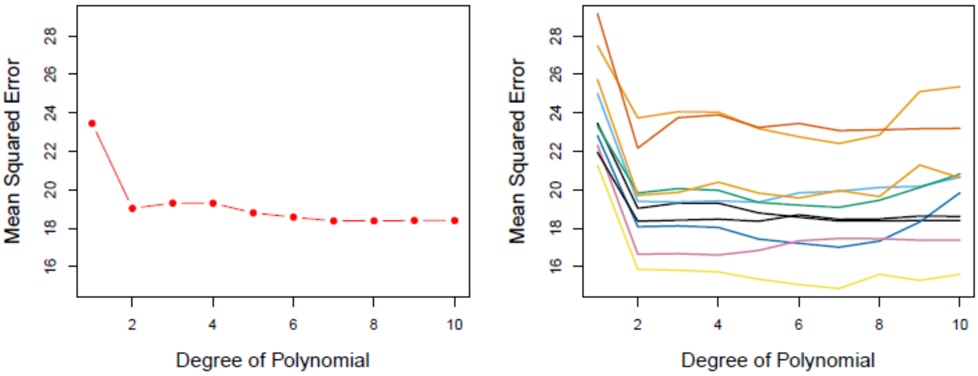

FIGURE 5.2. *The validation set approach was used on the* Auto *data set in order to estimate the test error that results from predicting* mpg *using polynomial functions of* horsepower. Left: *Validation error estimates for a single split into training and validation data sets.* Right: *The validation method was repeated ten times, each time using a different random split of the observations into a training set and a validation set. This illustrates the variability in the estimated test MSE that results from this approach.*

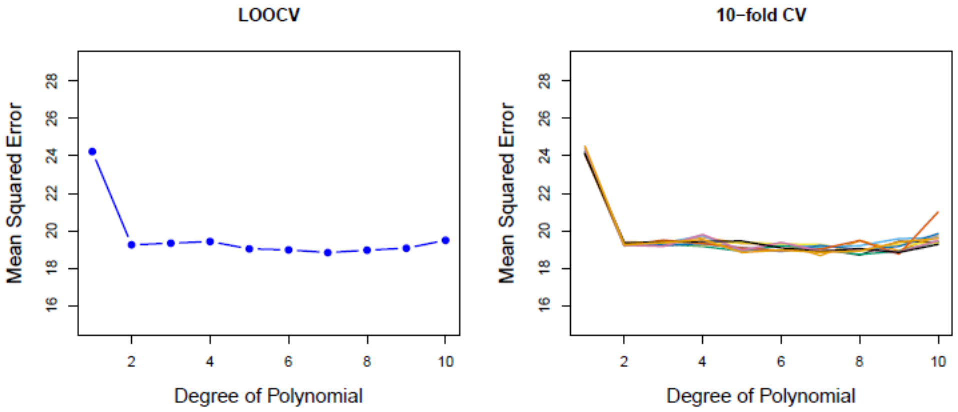

FIGURE 5.4. *Cross-validation was used on the* Auto *data set in order to estimate the test error that results from predicting* mpg *using polynomial functions of* horsepower. Left: *The LOOCV error curve.* Right: 10*-fold CV was run nine separate times, each with a different random split of the data into ten parts. The figure shows the nine slightly different CV error curves.*

ii. LOOCV?

Answer: Leave-one-out cross-validation (LOOCV) exhibits a reduced bias when compared to k-fold cross-validation because it utilizes nearly all the data points, making it almost unbiased. However, it introduces an element of the bias-variance trade-off, as LOOCV tends to have higher variance than k-fold cross-validation. This increase in variance occurs because, in LOOCV, the validation errors are highly correlated, with each pair of the n-1 fitted models differing by only 2 data points, as these models share n-2 out of the n-1 points used for each fit. In contrast, k-fold cross-validation significantly reduces this correlation, especially when k equals 5 or 10, for example.

To illustrate this with a k=10 example, each pair of the 10 fits shares approximately 80% of the data points, while with k=5, each pair of the 5 model fits shares roughly 60% of the points. Furthermore, k-fold cross-validation is computationally less demanding, as it necessitates only k model fits, in contrast to the n fits required for LOOCV. It's worth noting, as indicated in Equation 5.2 of the text, that there is an exception to this rule in the case of least squares linear or polynomial regression, where LOOCV requires the same computational effort as a single model fit.

4. Suppose that we use some statistical learning method to make a prediction for the response Y for a particular value of the predictor X. Carefully describe how we might estimate the standard deviation of our prediction.

Answer:

Bootstrap is a valuable statistical technique that enables us to estimate various properties of an estimator, such as the standard deviation. To estimate the standard deviation using bootstrap, we employ a process of repeatedly sampling the original data set with replacement, effectively creating numerous new data sets. In our specific case of estimating the standard deviation, the first step is to execute the bootstrap procedure a large number of times, often around 10,000 iterations. Subsequently, for each of these bootstrap samples, we fit a model and make predictions. The standard deviation of this collection of predictions is then employed to estimate the standard deviation of our predictor. It's important to note that the versatility of the bootstrap method extends beyond estimating standard deviation; it can be applied to assess other essential properties of an estimator, including the mean, confidence intervals, and prediction error, making it a powerful tool in statistical analysis.

Rest 5.5, 5.6 and 5.7 are in Python file.