

Chapter 5

3. *We know review k -fold cross-validation is implemented.*¹

(a) *Explain how k -fold cross-validation is implemented.*

You would implement k -fold cross-validation by taking the observations (n) and splitting them into k groups. You split them up randomly and the groups will not overlap; there is no union between the groups. Each group is a “validation set” for one MSE_i estimate, while the others will be a training set when they are not the validation set. You then find the mean of the MSE_i results, which will end up being the mean of the all the MSE estimates. There will end up being k MSE estimates.

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

i. *The validation set approach?*

k -fold cross-validation is harder to implement and to understand compared to the validation set approach. This is because you put the data into two sets in the validation set approach. But k -fold cross-validation will win out in a few ways; the validation set will have an error rate that is larger than the actual error rate when you fit the model and this estimate will have a higher variation than the k -fold cross validation approach will. This last one is because of the different observations that could end up being excluded in a validation set approach.

ii. *LOOCV?*

k -fold cross-validation has the advantage of easier to compute than LOOCV as well as having lower variance. But LOOCV has a smaller average bias than k -fold cross validation does.

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.*

We could use the bootstrap approach! When you are “bootstrapping”² you sample your observations repeatedly, with replacement. This might mean that you end up having the same observation ten times in your sample while another has not been sampled at all. You will do this a large number of times. Then you will get the MSE for each model. An equation from the book, equation 5.8, is below and can be useful:

$$SE_B(\hat{\alpha}) = \sqrt{\frac{1}{B-1} \sum_{r=1}^B \left(\hat{\alpha}^{*r} - \frac{1}{B} \sum_{r'=1}^B \hat{\alpha}^{*r'} \right)^2}$$

¹We talked about a lot of this problem during the beginning of class Wednesday *I think* (not explicitly).

²Can you say this?