



Homework 5
Patricia Hoffman, PhD.

- 1) In homework 3 problem 4 ridge regression was performed on the wine quality data set : [winequality-red.csv](#). Now use k-nearest neighbors to classify this data. Use cross validation to choose the best value for k. Round the results from the ridge regression to the nearest integer to form the classification with ridge regression. Using the best values for k (nearest neighbor) and lambda (ridge regression), compare and contrast the results of these two classification techniques.
- 2) Use k-nearest neighbors to classify the Iris data set. Compare the k-nearest neighbor results with the results obtained in class using the Naive Bayes Classifier.
- 3) Classify the wine quality data using Naive Bayes. Compare the results with the two methods described in problem 1 of this homework set. Think about why one of the methods used works better than the others.
- 4) Classify the sonar data using Naive Bayes. Compare the results with the methods used in class and with the last homework set. Give reasons for any discrepancies between the results for these methods. (Either in class or in homework, the following methods have been used on this data set: Trees, Linear Regression, Ridge Regression, an Ensemble Method, and now Naive Bayes.)
- 5) Run the code in the file KfirstNearestNeighbor.R Does KNN create a better model if the data is first scaled and normalized? What should be chosen as the best value for k and why? Now use KNN with cross validation on the mixtureSimData.data. What is the best value for k for this data?