Assignment 1 Answers

Name: Sam Robinson

1. Read WineQuality.pdf.
2. Use the RedWhiteWine.csv/arff file that is provided.  
   Note: If needed, remove the quality attribute, which you will not need for this assignment.
3. Build an experiment using Decision Trees.
   1. What is the percentage of correct classification results (using all attributes)?

Accuracy : 0.9792

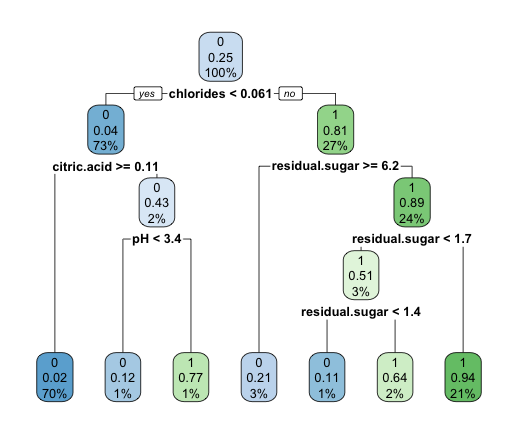
* 1. What is the percentage of correct classification results (using a subset of the attributes)?

Using: "alcohol", "residual.sugar", "pH", "chlorides", "citric.acid"  
Accuracy : 0.9461

* 1. What is the AUC of your model?

0.9355009

* 1. Visualize your decision tree



1. What is the best AUC that you can achieve?

I first tried to do a principal component analysis on the data to see how that interacted with logistic decision tree like the one we have here. It looks like that did a very poor job of improving the overall scores.

1. Which are the minimum number of attributes? Why?

Technically the minimum number of attributes to compare class to is 1. That doesn’t necessarily give you something predictive. The minimum number of variables I could get from pruning the tree and still having something predictive was 5. I achieved this by using the built in prune function in R.

