DataVizA Tutorial: Decision Trees

Department of Econometrics and Business Statistics, Monash University

Tutorial 11

Wine Data

1. Construct a decision tree using all data in *ExistingWines.rds* in the training set and predict the data in *NewWines.rds*.
2. Create a visual representation of the tree selected in Question 1. Use the default settings of `rpart.plot`
3. A wine has a Color Intensity of 3.9, Flavanoids of 1.8 and Proline of 800. What is your prediction for the market this wine is suited to? Use the output from the previous question to answer this rather than R.
4. What are the predicted probabilities that the wine in question 3 belongs to each class?
5. A wine has a Color Intensity of 2.4, Flavanoids of 1.2 and Proline of 700. What is your prediction for the market this wine is suited to? Use the output from the previous question to answer this rather than R.
6. What are the predicted probabilities that the wine in question 3 belongs to each class?
7. Using R, find predictions for the first ten wines in the *NewWines.rds*
8. Construct a different tree by requiring at least 15 training observations to be within each partition. Create a visual representation of this tree
9. Using R, find predictions for the first ten wines in the *NewWines.rds* using the tree obtained in the Question 8.
10. Split the data in *ExistingWines.rds* into a training sample (of roughly 70%) and a test sample (of roughly 30%).
11. Which tree is better for this data? How do these compare to the results from kNN and discriminant analysis from previous tutorials?