

Logistic Regression: Exercises/Assignments

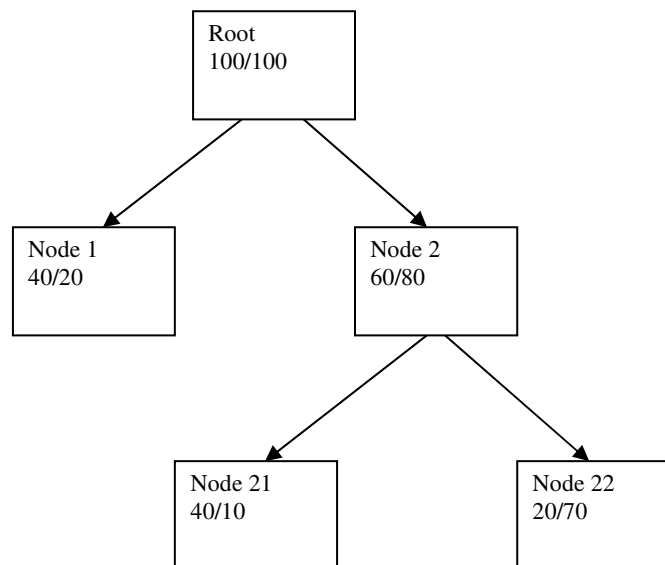
Exercise 1

Fit a logistic regression model to the stagec data and compare to the decision tree from Classification Trees Exercise 1.

The script `stagec_logistic.r` shows how to do this in R.

Exercise 2

Consider the following classification tree. For each node the number of +ve/-ve records is indicated.



For any $p \in [0,1]$ we classify the leaves of the tree as +ve/-ve according to whether or not the proportion of +ve records is $\geq p / < p$.

- For each value of $p \in [0,1]$ calculate the misclassification rate for this tree.
- Hence plot the ROC curve for this classifier.
- Suppose that the cost of a false +ve is c_+ and that the cost of a false -ve is c_- , what value of p minimizes the expected cost of using this classifier?

Assignment 4

Fit a logistic regression model to the spam data and compare to a decision tree. Use a training data set for fitting and model selection, and a test data set for comparison. Your comparison should include an ROC curve for the logistic regression model, on which is plotted a point for each of the three classification trees constructed in Assignment 3 (using different loss matrices).