Demonstration of use of classification trees

Use the `carseats` data, which is a part of the `ISLP` library. Convert `Sales` into a qualitative response. If `Sales` ≥ 8, then denote it as "High" else "Low".

1. Create a training set containing a random sample of 300 observations and a test set containing the remaining observations. Fit a tree to the training data with coded `Sales` (coded as high and low) as the response and all other variables as predictors. Interpret any one terminal node.

2. What is the training accuracy?

3. Predict the response using the test data and prepare the confusion matrix. What is the test accuracy?

4. Apply cross-validation to the training set in order to determine the optimal tree size with the weakest link pruning.

5. Produce a plot of the average cross-validated accuracy of the sequence of pruned trees vs effective αs (cost complexity parameters). Which effective α corresponds to the highest cross-validated classification accuracy (in the training set)?

6. Plot the number of nodes and tree-depths of the sequence of pruned trees vs effective αs.

7. Plot the training accuracy and test accuracy of the sequence of pruned trees vs effective αs.

8. Fit the optimally chosen tree to the train data.

9. Report the depth, number of nodes, training accuracy, and test accuracy of the optimal tree.

10. Compare the training and test accuracies between the pruned tree and the unpruned tree.

11. Fit a logistic regression to the data. Compute the training and test error.

12. Report the training and test error of

    (i) unpruned tree

    (ii) pruned tree

    (iii) logistic regression

    Compare the results.