

STAT 621 HOMEWORK 8

Due: Friday November 22

Problem 1: Recall the `Carseats` data set from the `ISLR` package. In our class example, `Sales` was converted to a binary variable with classes `High` and `Low` before fitting a classification tree. In this problem, you will use regression trees to predict the original numeric response `Sales`.

- a) Randomly split the data in half to create a training set and a testing set. Fit an unrestricted tree on the testing set. Plot the tree and discuss any interesting results. Report the MSE (labeled residual mean deviance in the summary output).
- b) Use the `cv.tree` function to estimate the optimal size tree. Construct a plot of tree size vs. deviance. Report the estimated optimal size.
- c) If necessary, prune the tree. Is the MSE reduced? Replot the pruned tree and comment.
- d) Use the fitted tree to predict `Sales` for the test set. Compute the prediction MSE.
- e) Reanalyze the data using the Bagging approach. Report the MSE and compare with your best tree above. What is the % variance explained? Use the `importance` function to determine which predictors are most important.
- f) Use a Random Forest to re-reanalyze the data. Report the MSE and compare with your best tree above. What is the % variance explained? Use the `importance` function to determine which predictors are most important.
- g) Summarize the results from all three procedures. What kind of overall conclusions can you draw about carseat sales?

Problem 2: Submit a 1-paragraph summary on the status of your project.