Problem Set 4 – Regression Trees and Random Forests

2016-04-26

Classification trees

- Load the data set *Carseats* from the package *ISLR* and also load the package *tree* which will be used for the exercise.
- Construct a binary variable *High*, which takes on a value of Yes if the Sales variable exceeds 8, and takes on a value of No otherwise. Hint: *ifelse()*
- Fit a classification tree to predict *High* using all other variables in the data set. Hint: *tree()*
- Plot the fitted tree.
- Determine the "best" tree by pruning / CV and prune to this tree on some training set. Hint: *cv.tree*
- Determine the quality of the predictions on the testing set. Hint: *predict()*

Regression Trees

- Fit a regression tree to the Boston data we had before in the class. The depedent variable is *medv*. Hint: package *MASS*
- Do this on a training set (50% of the sample) and then evaluate the predictions on the testing set.
- Plot the tree and interpret the results!

Bagging and Random Forests

- Repeat the excerise from above (fitting on training set + prediction on testing set) with bagging. Hint: *randomForest* from the package with the same name.
- Finally, fit a random forest on the training data and compare the model with all previous models.