Assignment #5: SVM and Random Forest

Use the data set OJ which is part of the ISLR package to answer the following questions.

(a) Create a training set containing a random sample of 800 observations, and a test set containing the remaining observations.

(b) Fit a **support vector classifier** to the training data using cost=0.01, with *Purchase* as the response and the other variables as predictors. Use the *summary()* function to produce summary statistics, and describe the results obtained.

(c) What are the training and test error rates?

(d) Use the *tune()* function to select an optimal cost. Consider values in range 0.01 to 10.

(e) Compute the training and test error rates using this new value for cost.

(f) Repeat (b-e) using **svm** with polynomial kernel with degree =2.

(g) Apply **random forest** to the dataset with *Purchase* as the response and the other variables as predictors and report the training and test error rates.

(h) Which method (random forest, support vector classifier or svm) is the best? Suggest the usage of these three methods for practice based on the above analyses.

Submit through ecampus link: eCampus -> Assignments

Deadline: April 14, Thursday *@11:59pm*