

Stat 897 Fall 2017 Data Analysis Assignment 9

Penn State

Due November 12, 2017

In this assignment we will use the OJ data found in the ISLR library.

- (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 the range 0.01 to 10.
- (e) Compute the training and test error rates using this new value for cost.
- (f) Repeat parts (b) through (e) using a support vector machine with a radial kernel. Use the default value for gamma.
- (g) Repeat parts (b) through (e) using a support vector machine with a polynomial kernel. Set `degree=2`.
- (h) Overall, which approach seems to give the best results on this data?