

## Assignment #7: SVM

### Problem 1

In this problem, you will use support vector approaches to predict whether a given car gets high or low gas mileage based on the **Auto** data set in the **ISLR** package.

- (a) Create a binary variable that takes on a 1 for cars with gas mileage above the median, and a 0 for cars with gas mileage below the median. Use this variable as response in the following analysis.
- (b) Fit a support vector classifier to the data with various values of **cost**, to predict whether a car gets high or low gas mileage. Report the cross-validation errors associated with different values of this parameter. Comment on your results.
- (c) Now repeat (b), this time using SVMs with radial and polynomial kernels, with different values of **gamma**, **degree** and **cost**. Comment on your results.

### Problem 2

This problem uses the **OJ** data set in the **ISLR** package.

- (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 value 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 **tune()** function to select an optimal **cost** and **gamma**.
- (g) Repeat parts (b) through (e) using a support vector machine with a polynomial kernel. Set **degree=2**. Use the **tune()** function to select an optimal **cost**.
- (h) Overall, which approach seems to give the best results on this data?

*Submit through link: eCampus -> Assignments*

*Deadline: Nov 14, Tuesday @11:59pm*