Problem Set 2: Support Vector Machines

BUAD 5082 – Spring 2019

1. Objectives

The purpose of this problem set is to provide you with an opportunity to practice the kinds of skills that I expect you to be able to perform on an exam.

2. What You Will Need

- Access to a Windows computer with a recent version of R installed.

3. Solutions

Solutions to these problems will be posted several days after this Problem Set is posted.

4. Tasks:

This Section involves the OJ data set which is part of the ISLR package.

a) Use the following code to create a set of indices containing a random sample of 800 integers representing the training subset of set OJ, and a set of test indices representing the remaining observations:

```
set.seed(5082)
n = dim(OJ)[1]
train_inds = sample(1:n,800)
test_inds = (1:n)[-train_inds]
```

b) Fit a support vector classifier to the training data using cost=1, with Purchase as the response and the other variables as predictors. Be sure to scale the predictors. Use the summary() function to produce summary statistics, and describe the results obtained.

c) Compute and display the training and test error rates?

d) Use the `tune()` function to select an optimal cost. Use the default setting for gamma and consider the following cost values: c(0.01, 0.05, 0.1, 0.5, 1, 5).

e) Display a summary of the best model and compute and display the training and test error rates using this best model.

f) Repeat parts (d) and (e) using a support vector machine with a radial kernel. Search for the cost and gamma parameters that produce the smallest test MSE. Use a search grid composed of:

   a. Costs: c(0.01, 0.05, 0.1, 0.5, 1, 5)
   b. Gammas : c(0.001, 0.01, 1, 3, 5)

g) Repeat parts (d) through (e) using a support vector machine with a polynomial kernel. Search for the cost and degree parameters that produce the smallest test MSE. Use a search grid composed of:
    a. Costs: c(0.01, 0.05, 0.1, 0.5, 1, 5)
    b. Degree : c(2, 3, 4, 5)

h) Overall, which approach seems to give the best results on this data?