ISYE 6501 – Introduction to Analytics Modeling

Week 1 Homework

**Question 1**

Describe a situation or problem from your job, everyday life, current events, etc., for which a classification model would be appropriate. List some (up to 5) predictors that you might use.

**Response –**

The classification model can be useful to find out fraudulent credit card transactions of a customer. Possible predictors could be dollar amount, location, type of purchase, time of purchase, frequency of the item purchase.

**Question 2.1**

Using the support vector machine function ksvm contained in the R package kernlab, find a good classifier for this data. Show the equation of your classifier, and how well it classifies the data points in the full data set. (Don’t worry about test/validation data yet; we’ll cover that topic soon.)

**Response –**

Please see the code attached.

The classifier equation is: (-0.00046)A1 + (-0.014)A2 + (-0.0081)A3 + (-0.01)A8 + (-0.501)A9 + (-0.0014)A10 + (-0.00129)A11 + (-0.00026)A12 + (-0.206)A14 + (558.33)A15 + 0.0815 = 0

The above classifier has 86.4% accuracy in classifying the data.

Different values of C ranging from 0.0001 to 1000 are tried. C value of 0.0001 gave a 54% while C values from 1 to 1000 gave ~86% accuracy and didn’t vary much.

**Question 2.2**

Using the k-nearest-neighbors classification function kknn contained in the R kknn package, suggest a good value of k, and show how well it classifies that data points in the full data set. Don’t forget to scale the data (scale=TRUE in kknn).

**Response –**

Please see the code attached.

Various k values ranging from 2 to 10 are used to check which k value predicts the responsible variable with more accuracy. The accuracy increases from 81.515 to 85 .16% as k value increases from 2 to 5 and the accuracy decreases again when k > 5. Hence, 5 is a good value for k with an accuracy of 85.16%