**Homework Assignment 1 by Haritha Pulletikurti**

**Question 2.1**

**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.**

# Answer:

Below are the two examples:

# Budget Classification: We can use a classification model to build a classifier to categorize the monthly or yearly spending into Expensive, Inexpensive categories.

The Factors or predictors we can have for this may include: Grocery, School Fees, Utilities, House/Mortgage payments etc.

# Job Applicant:

The Predictors include “Highest Level of Education”, “Years of Experience”, “Awards”, “Communication Skills”, “Required skillset for the job”, “Age” etc. and the Classification Results are categorized as either “Success” or “Failure”.

**Question 2.2**

# Question 1 : Find a good C Value for KSVM?

# Analysis:

# Please see the attached R report for further details. I am providing some additional analysis in this document.

* For the C Values: 0.1, 1, 10 ,100, 1000 and 10000

Some of the results of the ksvm function are shown in the following table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| C Value (Ordered by Highest Accuracy ) | Accuracy Percentage | Error | Variance | Bais | No.of Support Vectors |
| 100 | 86.39144 | 0.1360856 | Better than all other cases . less high than other cases | Better than all other cases.Less low than other cases. | 189 |
| 10 | 86.39144 | 0.1360856 | high | low | 190 |
| 1 | 86.39144 | 0.1360856 | highest | lowest | 190 |
| 0.1 | 86.39144 | 0.1360856 | highest | lowest | 197 |
| 10000 | 86.23853 | 0.1376147 | low | high | 284 |
| 1000 | 86.23853 | 0.1376147 | low | high | 275 |

# As C value increases, flexibility increases, margin increases, number of violations that can occur also increase. Also, number of Support Vectors Increase, and this result in higher bais but low variance

# As C value decreases, Flexibility decreases, margin decreases and the number if violations that can occur is less. Also, number of support vectors decrease and this result in lower bais and high variance.

# According to research, As we increase the flexibility of the classifier, bias initially tends to decrease faster than the variance increases. This makes the error decline. However, at somepoint, increasing the flexibility has little impact on bais and significantly increases the variance.

# The above listed data shows that C=100 with lowest number of support vectors will be a better classifier than the rest according to the above listed information.

# Question 2:

# Answer:

# I tried the same c=100 with another kernel called Polydot. The result for the classifier was same as “Vanilladot” kernel. Please see the implantation file for more details.

# Question3 : KKNN implantation.

# It is determined that for k =12 and k = 15, the classifier is optimal with highest accuracy of 85.321110.

Based on the plot shown in the implementation file, the accuracy of the prediction decreased as k >15.

# Hence the optimal value of k = 12.

# Please see the attached implantation details of the above analysis.