**Using Credit Approval Data to find a good classifier of loan applicant**

**Shafagh Yazdani**

***Data***

*The files credit\_card\_data.txt (without headers) and credit\_card\_data-headers.txt (with headers) contain a dataset with 654 data points, 6 continuous and 4 binary predictor variables.  It has anonymized credit card applications with a binary response variable (last column) indicating if the application was positive or negative. The dataset is the “Credit Approval Data Set” from the UCI Machine Learning Repository (*[*https://archive.ics.uci.edu/ml/datasets/Credit+Approval*](https://archive.ics.uci.edu/ml/datasets/Credit+Approval)*) without the categorical variables and without data points that have missing values.*

***Project Overview***

Using Credit Approval Data Set (654 data points, predictor variables: income, credit history,age, family size, assets, liabilities ,) we run the ksvm/knn classification model(in the R package kernlab) to find a good classifier to find which loan applicant will pay the loan.

Our KSVM model with coefficients is:

( -0.0001500738) X1 + (-0.0014818294 ) X2 + 0.0014083130 X3 + 0.0072863886 X4 + 0.9916470037 X5 +( -0.0044661236) X6 + 0.0071482899 X7 + ( -0.0005468386 )X8 +(-0.0016930578 )X9 + 0.1054824270 X10 +(-0.08198854)

The best classifier is the one that have maximum margin from both class.

Based on the prediction of Response in our ksvm model the best C value is the one that have highest accuracy of response prediction.

Part 1

|  |  |
| --- | --- |
| C Value | Model Accuracy |
| C=0.00001 | 0.5474006 |
| C=0.001 | 0.5474006 |
| C=0.05 | 0.8639144 |
| C=0.01 | 0.8639144 |
| C= 10 | 0.8639144 |

Before running the KKNN model we split the data to two-part training and test data. And we train our model on training data and then test the model on test data and predict the response on test data. Due to possible overfitting issue we need to test our model on test data.

Based on the prediction of Response in our kknn model the best K value is the one that have highest accuracy of response prediction on our test data

|  |  |
| --- | --- |
| K Value | Model Accuracy |
| K= 2 | 0.7738693 |
| K=4 | 0.7738693 |
| K=8 | 0.8341709 |
| K=12 | 0.839196 |
| K=19 | 0.8442211 |