LI LIPING (A0186040M)

BT5152: Decision Making Technology for Business

Semester 1, AY 2018/19

Assignment 1

#the code is in the rmarkdown file, this PDF just concludes all the train and test accuracy.

1. (6 marks) Model the training data "loan\_train.csv" using KNN, Naïve Bayes, C50 decision tree decision tree receptively. Report training accuracies and test accuracies on the training dataset "loan\_train.csv" and test dataset "loan\_test.csv" respectively.

* Remember to scale your numerical variables properly and convert categorical variables by OneHot for KNN.

|  |  |  |
| --- | --- | --- |
|  | train accuracy | test accuracy |
| KNN | 0.9999638 | 0.7169888 |
| Naïve Bayes | 0.8046055 | 0.8012397 |
| C50 decision tree | 0.8184166 | 0.8067857 |

2. (6 marks) Now we practice rpart package. In order to avoid over fitting, prune the decision tree using three **pre-pruning** methods, and **post-pruning by best complexity parameter**. Compare the accuracies of fully-grown tree and 4 trees (both on training set and testing set) of the decision tree classifier. Discuss which tree gives you the best prediction results on the test set.

* Before pruning (the fully-grown tree in this assignment), please set cp= 1e-05 (0.00001).
* For the 3 pre-pruning, try minsplit = 800, minbucket = 200, and maxdepth = 3.
* **This bullet is not a requirement for this assignment. You are encouraged to try other pre-pruning parameters or change cp before pruning to understand more about how pruning affect the accuracy on the training set and test set.**

|  |  |  |
| --- | --- | --- |
|  | train accuracy | test accuracy |
| fully grown | 0.8534156 | 0.7713074 |
| pre1 | 0.8163051 | 0.8062964 |
| pre2 | 0.8166132 | 0.8065411 |
| pre3 | 0.8152901 | 0.8065411 |
| pruned | 0.8168669 | 0.8062964 |

Based on the train and test accuracy shown in the table above, we can see that pre-pruning2 (with control minbucket=200) and pre-pruning3 (with control maxdepth=3) give the best prediction results on the test set, which is 0.8065411