

RMSC 4002 Assignment 3 1st term 2017/2018

We shall use the same dataset “credit.csv” in assignment 2.

Question 1 (CTREE)

- (a) Read in credit.csv and save it in *d*. Use the last six digits of your birth date as the random seed, (For example, if your birth date is Dec. 10th, 1996, the random seed is 961210), randomly sample 580 records from *d* as the training dataset and save it in *d1*. Save the other records in *d2* as the testing dataset.
- (b) Using the training dataset *d1*, build a classification tree of Result with other variables. Using the default option in *rpart()* probably gives a very complicated tree. Therefore we should add in the option *control=rpart.control(maxdepth=3)* inside the *rpart()* function. (See *help(rpart)* and *help(rpart.control)* for more details).
- (c) Plot the tree with *use.n=T* and print the result. Write down the classification rules from the output. Compute the confidence, support and capture for each rule.
- (d) Produce the classification table and compute the training error rate.
- (e) Apply the classification rules in (d) using the testing dataset *d2* and produce the corresponding classification table. Compute the testing error rate.

Question 2 (ANN)

- (a) Using the same dataset *d1* and *d2* in question 1 (a) as the training dataset and testing dataset. Fit an improved version *ann()* function with *size=6*, *linout=T*, *maxit=500* and *try=25* and save the output to *ann6*.
- (b) Repeat part (a) using *size=7*, *8* and *9*. Save the result to *ann7*, *ann8* and *ann9* respectively.
- (c) Compare the final value in *ann6*, *ann7*, *ann8* and *ann9*. Choose the best (smallest) one and produce the classification table for the training dataset *d1*. Compute the training error rate.
- (d) Use the best ANN model in part (c), produce the classification table for the testing dataset *d2* and hence compute the testing error rate.
- (e) Compare and comment on these results obtained with the results in Question 1.

Submit the files: asg3-1.doc, asg3-2.doc, and asg3-1.r, asg3-2.r containing answers and fully commented R commands via eLearning system on or before December 4, 2017 and put a hard copy in the drop-box on or before December 5, 2017.

Reminder: Choose **one** member in your group to submit your final project on or before **December 1, 2017** and put a hard copy of your report in the drop-box on or before **December 2, 2017**.