Department of Mechanical Engineering
ME 781: Engineering Data Mining and Applications

Assignment-3

Mandatory reading:

1. Confusion Matrix: http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/
2. ROC and AUC: https://stats.stackexchange.com/questions/105501/understanding-roc-curve
3. Understand the ***confusionMatrix*** function from package "caret", and how to use it.
4. What is meant by ***accuracy***, ***sensitivity*** and ***specificity***? What are the alternate terms by which sensitivity and specificity are known?
5. What is ***kappa*** and what significance does have in the context of classification?

With reference to the given data set **a3-data-set.csv** perform the following:

1. Read in the data set into a data frame. It is a classification data set. Review the data.
2. The requirement is to run various models (list given below) and note down the various classification metrics in each case (metrics Table given below)
3. List of models to be run on the data set:
    - K Nearest Neighbours (with 3 and 5 neighbours)
    - SVM Linear
    - SVM Radial
    - Random Forest
    - Neural Network with 3 nodes in the hidden layer
    - Neural Network with 5 nodes in the hidden layer
4. In each of the above cases, call the ***confusionMatrix*** function from the "caret" package and summarize the metrics as shown in the following Table.

	Knn.3	Knn.5	svmLinear	svmRadial	RForest	Nnet.3	Nnet.5
Accuracy							
Kappa							
Sensitivity							
Class 1							
Class 2							
Class 3							
Class 4							
Specificity							
Class 1							
Class 2							
Class 3							
Class 4							

5. Based on the classification metrics captured in the Table, decide which method best models the data.
6. Repeat the above exercise using the train and trainControl functions of the caret package to implement 10-fold cross validation. Based on this, decide which method best models the data.

7. Optional: Plot the class boundaries for each of the methods. (Hint: see the region plotting code given in ***neural-network-classification.txt***

Note:

- Submit your answers to the assignment submission point in Moodle.
- This is one of a set of 5 Problems that you will have to solve. Submission of all assignments will fetch you max 5 marks.
- The Test scheduled on Nov-8-2017 will assume you have done this assignment.