BUAN 6341
APPLIED MACHINE LEARNING
ASSIGNMENT 3
Due date: November 4, 11:59 pm

In this assignment, you will be using the following two learning algorithms:
1.   Artificial neural networks (ANN)
2.   K Nearest Neighbors

You must use R or Python for this assignment. You **can** use **any** publicly available R or Python library/package. You can even use H2O.

For the following tasks, use the same two datasets that you used in assignment 2.

Tasks:

1.   Download and use any neural networks package to classify your classification problems. Experiment with number of layers and number of nodes, activation functions (sigmoid, tanh, etc.), and may be a couple of other parameters.

2.   Download and use any KNN package to classify your classification problems. Experiment with number of neighbors. You can use any distance metric appropriate to your problem. Just be clear to explain why you used the metric that you used.

Deliverables:
You are required to submit the following:
•   Your code file(s)
•   A readme file explaining how to run your code.
•   Report (must not exceed 8 pages total)
•   Any supporting files (data sets, etc. If the data sets are too large, submit the url)

Your report should be both thorough and concise and contain at the very least the following:
•   Small description of the data sets and the classification problem.

•   Error rates (train and test) for the two algorithms on your two data sets. Plot various types of learning curves that you can think of (e.g. but not limited to – error rates vs. train data size, error rates vs. clock time to train/test, etc.).

•   Performance comparisons (learning curves, confusion matrices, etc.) of various functions/parameters for the algorithms (e.g. ANN number of layers, nodes, etc., number of neighbors in KNN, etc.) on both the data sets.

•   Comparisons of the two learning algorithms using the two data sets.

- Does any of these algorithms perform better than the three algorithms used in assignment 2? What is your observation about how these 5 algorithms performed on the two data sets (ranking of algorithms for these data sets)?

Be creative and think of various experiments that you can come up with for this assignment. You need to give clear description of all your experiments and analysis. Why did you get the results that you did? Compare and contrast the algorithms. What additional things can you do to get better results? Would cross validation help? If yes then why didn't you implement it? How did you pick various parameters, and how did they compare? Which algorithm performed the best and why? How did you define best? Think of as many questions as you can! This assignment will take time. So get started on it today!

Grading:
Total weightage: 10% of final grade

Breakdown:

Code: 0 points (Code should execute and produce the results presented in the report with minimum effort – We will run the code and if it doesn't run or has errors, points will be deducted from the report).

Report: 100 points
Points will be awarded based not only on how good your results are, but also on how well you describe them as well as underlying experimentation. Any plots without explanation = 0 points. Similarly, explanation without plots = 0 points. Keep in mind that you are graded on your analysis and description, as well as creativity. Have fun!