

BUAN 6341
APPLIED MACHINE LEARNING
Assignment 2
Due date: July 24, 11:59 pm

In this assignment, you will be using the following learning algorithm:

1. Artificial neural networks (ANN)

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

Tasks:

1. You are to use the data set from assignment 1 for this assignment.
 - Convert the data set from assignment 1 to a binary classification problem by thresholding the output to a class label. Use this transformed data set for all the tasks and experiments in this assignment.Divide your data set in train and test sets.
2. 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.
3. Implement and use cross validation for your dataset with the above algorithm.

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 6 pages total)
- Any supporting files (data set, etc. If the data set is too large, submit the url)

Your report should be both thorough and concise and contain at the very least the following:

- Error rates (train and test) for the neural network algorithm on your data set. 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 algorithm (e.g. ANN number of layers, nodes, etc.).

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? Were these results good or bad because of the specific data set? Compare and contrast the different algorithms. What additional things can you do to get better results? How did cross validation help? How

did you pick various parameters, and how did they compare? Think of as many questions as you can!
This assignment will take time. So get started on it today!

Grading:

Total weightage: 15% 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!