Assignment - Final Project

Objective: to create and improve an artificial neural network (ANN) using techniques we have learned in this course. This is your project. Find something interesting to work on, within the constraints below.

Data sources: You may choose a dataset of your choice from either <https://www.kaggle.com/datasets> or <https://archive.ics.uci.edu/ml/datasets.html>. If you wish to use another datasource, please get approval from your professor first.

Data characteristics: The data should have at least 5 usable attributes or columns (I.e. you will use them in your ANN model), and have at least 4000 instances or records. Exceptions from these rules must be approved by your professor.

Model type: You are to develop an ANN model(s) appropriate for your data and objective.

Deliverable 1 – A one page paper (Please justify all your choices):

1. Your objective should be stated in written form. What are you trying to accomplish? Predicting a number? Classifying? Your objective must reference the context of the problem, specifically.

*Good example:*

*"The aim is to distinguish between the presence and absence of cardiac arrhythmia and to classify it in one of the 16 groups. Class 01 refers to 'normal' ECG classes 02 to 15 refers to different classes of arrhythmia and class 16 refers to the rest of unclassified ones. For the time being, there exists a computer program that makes such a classification. However there are differences between the cardiolog's and the programs classification. Taking the cardiolog's as a gold standard we aim to minimise this difference by means of machine learning tools."* [*https://archive.ics.uci.edu/ml/datasets/Arrhythmia*](https://archive.ics.uci.edu/ml/datasets/Arrhythmia)

*Bad example:* “My objective is to classify mud.”

1. Characteristics of your data choice and justification
2. The type of model you intend to develop.

Deliverable 2 – A final paper (Please justify all your choices):

1. Your objective should be stated in written form. What are you trying to accomplish? Predicting a number? Classifying? Your objective must reference the context of the problem, specifically.
2. Your final ANN model, in code.
3. Your final model and training algorithm, in words.
4. Your experimental plan for arriving at the final model.
5. How long it took to run all the models in your experimental plan.
6. An explanation of the input variables and any preprocessing steps you took.
7. An explanation of your metrics and justification for your choice.
8. An explanation of your method to validate the model.
9. Your results in terms of appropriate metrics for the objective and problem.
10. A discussion and/or justification for how you used/didn’t use all of the following:
    1. selection of the optimum number of units
    2. type of network (feedforward, recurrent, backpropagation, and etc.),
    3. type of training (supervised, unsupervised),
    4. proportion of training and testing data sets (70:30, 80:20, and etc.),
    5. number of input and output units (usually application dependent),
    6. number and size of hidden layers (2N+1, experimental),
    7. number of repetitions during training (epoch),
    8. choice of activation function (sigmoid, linear, Tanh, ReLU, and etc.) and
    9. size of data set (number of records)
    10. learning rate
    11. momentum.
11. Discussion and further work.

Charts and graphs should be used as appropriate.

This is an Honor code Category C assignment.

Category C – This is an individual assignment. You may work with others or receive help from a tutor on this assignment. You must, however, turn in your own paper. You may not divide the work with others or copy another student’s paper or work; it is an honor code offense to do so.