CS 422/622
INTRODUCTION TO MACHINE LEARNING

FALL 2023
Assignment #6

Due Date/Time: **12/08/2023 @ 11:59PM**
Total Points: **100**

CS 422 students may complete this assignment individually or in teams of two; CS 622 students are to complete it individually.

Description:

➢ For this assignment, you will implement ***multi-layer perceptron (MLP) regressor*** or ***MLP classifier*** in Python or MATLAB to solve a ***regression problem*** or a ***classification problem*** of your choosing. You may use any and all built-in methods.

➢ Train and test your model with a dataset of your choosing that meets the following criteria:

   o Number of input features: 3+
   o Input characteristics: Continuous real-valued
   o Output characteristics: Continuous real-valued for regression OR categorical for classification

➢ You may use one of the datasets you used in Assignment 3 or Assignment 4; or if you'd like, you may use one of the following sources to find a dataset:

   o University of California, Irvine Machine Learning Repository https://archive.ics.uci.edu/ml/index.php
   o Kaggle https://www.kaggle.com/
   o Awesome Public Datasets https://github.com/awesomedata/awesome-public-datasets
   o Google Dataset Search Engine https://datasetsearch.research.google.com/
   o Microsoft Research Open Data https://msropendata.com/
   o U.S. Government's Open Data https://www.data.gov/
   o Registry of Research Data Repositories https://www.re3data.org/
   o CMU Libraries https://guides.library.cmu.edu/machine-learning/datasets

➢ Use 80% of the dataset for training and 20% for testing your model.

➢ Experiment with ***three (3)*** different combinations of model parameters (e.g., number of hidden layers, number of neurons at each hidden layer, activation function, learning rate, etc.).

➢ Summarize your approach and results in a report that includes at least the following:

   o Name(s) of the student(s) completing the assignment

   o The dataset you used, its source and characteristics.

   o The data preprocessing steps you took (if any).

- o A table showing relevant evaluation metrics (MSE, MAE, R^2 for regression OR accuracy, sensitivity, specificity, f1 score, log loss for classification) for the training dataset with the three different combinations of model parameters you experimented with.

- o A table showing relevant evaluation metrics (MSE, MAE, R^2 for regression OR accuracy, sensitivity, specificity, f1 score, log loss for classification) for the test dataset with the three different combinations of model parameters you experimented with.

- o Any additional details you would like to include.

- ➢ Submit your report along with your dataset and source code. Feel free to include your code in the report, but you also need to submit your source code files (.py or .m) and your dataset separately, so that your results can be replicated for grading.

Submission Instructions:

Compress all files and submit it through Canvas as a **.zip** file.

If you are completing the assignment as a team of two, only one of the team members needs to submit the assignment.

I will set Canvas to allow unlimited number of submissions and will only grade the last submission. So, please do not wait until the last minute to submit as you can always submit an updated version.

Submission Instructions:

Compress all files and submit it through Canvas as a **.zip** file.

If you are completing the assignment as a team of two, only one of the team members needs to submit the assignment.

I will set Canvas to allow unlimited number of submissions and will only grade the last submission. So, please do not wait until the last minute to submit as you can always submit an updated version.