HPE DSI 311 – Introduction to Machine Learning – Spring 2024
Homework Assignment #3
Due Monday, July 1st, 11:59 pm (Central)

Your assignment is to create a Jupyter notebook that demonstrates how to do the following (use methods discussed in the class materials shared so far):

Load the dataset in the file named winequality_white.csv and set up a classification problem: predicting the quality value (a single variable y with seven classes labeled 3, 4, 5, …, 9) based on the values of **all** the other variables (acidity, alcohol, pH, etc.), as in H/W assignment #2 . The goal of this assignment is to see how an MLPClassifier performs compared to a more classical ML model.

1. Train three MLPClassifier models (via cross-validation on the training set) using different combinations of architecture choices (e.g., number of layers, # of neurons per layer, activation function). Report and discuss statistics for the validation scoring method of your choice; (**6 points**)
2. Pick the best performing model from Step 1. Use cross-validation on the same training set to study the impact on the model's performance when varying three different hyperparameters for the optimizer (e.g., solver, epoch, learning rate); Report and discuss statistics for the same validation scoring method as in Step 1; (**6 points**)
3. Pick the best performing model from Step 2. Test its performance using three different scoring methods. Discuss in detail your results, especially if there seems to be over/underfitting. (**5 points**)
4. Train and tune a new classifier that is not a neural network (you can use one from H/W assignment #2 if you want), using the training set from Step 1. Test this new model's performance on the corresponding test set and compare to the MLPClassifier test results from Step 3. Report and discuss statistics for three different scoring methods. (**5 points**)

What to submit: Please name your h/w submission as follows:
311_lastName_firstName_assignmentNumber.ipynb

How to submit: Please submit your homework in Moodle.