Assignment No. 3
EECS 658
Introduction to Machine Learning
Due: 11:59 PM, Thursday, September 30, 2021
Submit deliverables in a single zip file to BlackBoard
Name of the zip file: FirstnameLastname_Assignment3 (with your first and last name)
Name of the Assignment folder within the zip file: FirstnameLastname_Assignment3

Deliverables:

1. Copy of Rubric3.docx with your name and ID filled out (do not submit a PDF)
2. Python source code for CompareMLModelsV2
3. Screen print showing the successful execution of CompareMLModelsV2
4. Answers to the following questions for CompareMLModelsV2:
    a. Based on accuracy which model is the best one?
    b. For each of the 11 other models, explain why you think it does not perform as well as the best one.
5. Python source code for dbn.py
6. Screen print showing the successful execution of dbn.py
7. Answers to the following questions about dbn.py:
    a. Does the program use k-fold cross-validation?
    b. What percentage of the data set was used to train the DBN model?
    c. How many samples are in the test set?
    d. How many samples are in the training set?
    e. How many features are in test set?
    f. How many features are in the training set?
    g. How many classes are there?
    h. List the classes.

Assignment:

- This assignment has two parts:
- Part 1: Expand our comparison ML classifiers to include SVM, Decision Tree, Random Forest, ExtraTrees, and Neural Network.
    - Enhance your CompareMLModels program and call it CompareMLModelsV2 so that it includes the SVM, Decision Tree, Random Forest, ExtraTrees, and Neural Network.
    - It should now do the following:
        - Uses 2-fold cross-validation to produce a test set of 150 samples of the iris data set with the following ML models:
            - Naïve Baysian (NBClassifier)
            - Linear regression (LinearRegression)
            - Polynomial of degree 2 regression (LinearRegression)
            - Polynomial of degree 3 regression (LinearRegression)
            - kNN (KNeighborsClassifier)
            - LDA (LinearDiscriminantAnalysis)
            - QDA (QuadraticDiscriminantAnalysis)
            - SVM (svm.LinearSVC)

        - Decision Tree (DecisionTreeClassifier)
        - Random Forest (RandomForestClassifier)
        - ExtraTrees (ExtraTreesClassifier)
        - NN (neural_network.MLPClassifier)
    - For each of the 12 models the program should display (with a label before each model's display indicating which model the results are for):
        - Confusion matrix
        - Accuracy metric
    - If the values in your confusion matrices do not add up to 150, then you did something wrong.
- Part 2: Implement the deep learning DBN example at:
    - https://github.com/albertbup/deep-belief-network/blob/master/README.md
    - Name the program dbn.py
    - The code has two imports from SupervisedDBNClassification. Use the one "from dbn import SupervisedDBNClassification" and comment out the other one. Note: The sample code uses "from dbn.tensorflow import SupervisedDBNClassification" and has "from dbn import SupervisedDBNClassification" commented out.
    - More details regarding the program can be found at: https://medium.com/analytics-army/deep-belief-networks-an-introduction-1d52bb867a25

Remember:

- Your Programming Assignments are individual-effort.
- You can brainstorm with other students and help them work through problems in their programs, but everyone should have their own unique assignment programs.