Assignment No. 1
EECS 658
Introduction to Machine Learning
Due: 11:59 PM, Thursday, September 2, 2021
Submit deliverables in a single zip file to BlackBoard
Name of the zip file: FirstnameLastname_Assignment1 (with your first and last name)
Name of the Assignment folder within the zip file: FirstnameLastname_Assignment1

Deliverables:
1. Copy of Rubric1.docx with your name and ID filled out (do not submit a PDF)
2. Python source code for CheckVersions
3. Screen print showing the successful execution of CheckVersions
4. Python source code for NBClassifier
5. Screen print showing the successful execution of NBClassifier
6. Answers to the following questions:
    a. How many samples in NBClassifer were in the training set?
    b. How many samples in NBClassifer were in the test set?
    c. Using the confusion matrix, manually calculate the Accuracy value. Does it match the value calculated by your program? If not, why? (Manually includes using a spreadsheet).
    d. Using the confusion matrix, manually calculate the Precision values for each iris variety. Do they match the values calculated by your program? If not, why?
    e. Using the confusion matrix, manually calculate the Recall values for each iris variety. Do they match the values calculated by your program? If not, why?
    f. Using the confusion matrix, manually calculate the F1 values for each iris variety. Do they match the values calculated by your program? If not, why?

Assignment:
- Install Python on your system if it is not already.
    o See "Python for Windows Primer" on BlackBoard (under Assignment 1) for help on Windows
    o For help on Linux, see:
        ▪ https://wiki.ittc.ku.edu/ittc_wiki/index.php/EECS168:SSH_Instructions
        ▪ Virtual Box: https://www.virtualbox.org/wiki/Downloads
        ▪ Ubuntu install: https://ubuntu.com/download/desktop
    o See "Beginner's Python Cheat Sheet" on BlackBoard (under Assignment 1) for help with Python.
- Install the following Python libraries.
    o scipy
    o numpy
    o pandas
    o sklearn

- The scipy installation page provides excellent instructions for installing the above libraries on multiple different platforms, such as Linux, mac OS X and Windows. If you have any doubts or questions, refer to this guide, it has been followed by thousands of people.
- To verify you have installed Python and the SciPy libraries write a Python program called CheckVersions that 1) prints out the versions of Python, scipy, numpy, pandas, and sklearn and 2) prints out "Hello World!"
    - Hint: use this code for part 1):
        ```
        # Python version
        import sys
        print('Python: {}'.format(sys.version))
        # scipy
        import scipy
        print('scipy: {}'.format(scipy.__version__))
        # numpy
        import numpy
        print('numpy: {}'.format(numpy.__version__))
        # pandas
        import pandas
        print('pandas: {}'.format(pandas.__version__))
        # scikit-learn
        import sklearn
        print('sklearn: {}'.format(sklearn.__version__))
        ```
- Write a Python program called NBClassifier that does the following:
    - Loads the iris data set (located in iris.csv file in the BlackBoard Assignment 1 folder)
    - Creates a training set with half of the 150 samples and a test set with the rest.
    - Classifies the iris data set using the Python built-in Naïve Bayesian classifier, GaussianNB.
    - Prints out the overall accuracy of the classifier.
    - Prints out the confusion matrix.
    - Prints out the P, R, and F1 score for each of the 3 varieties of iris.
    - You may (and probably should) use the Python built-in programs.

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.