Fall 2020: CSCI 4/5588 Programming Assignment #2

DUE: Monday, Oct 19, 2020 (**Softcopy** @2 PM).

Total Marks = 100

Instructions

- All work must be your own (other than the instructor provided codes and hints to be used). You are NOT to work in teams on this assignment.
- Format: Your solution must be typed. Submit a single compressed file (via Moodle), preferably a doc or a pdf file. Name it as PA2_<Your_name>. Provide the hardcopy in the class.
- The top/cover page of the report should have the title, "Fall 2020: CSCI 4/5588, ML-II Programming Assignment #2". Then your, "Name:_____ and ID:_____

Description

This programming assignment is to train, and then compare performances of **5 different classifiers** of your choice **including SVM** from Weka (or, Scikit learn; or, LibSVM for SVM) to predict *coronary heart disease* (CHD) candidate.

- Use **(a)** 05-fold cross validation (5 FCV) and **(b)** 10-fold cross validation (10 FCV), to train and measure the performance of a classifier.
- You are welcome to generate any additional useful feature from the given dataset to be used as input feature.

Data

- Information about the dataset:
  - http://www-stat.stanford.edu/~tibs/ElemStatLearn/datasets/SAheart.info
  - Dataset: http://www-stat.stanford.edu/~tibs/ElemStatLearn/datasets/SAheart.data
- Check Moodle for a copy of the datasets and related information.

Reporting

- Describe each classifier in 5 to 10 sentences and provide an appropriate reference(s).
- Describe the parameters/hyper-parameters that you have chosen to train the classifiers.
- Define and describe the following terms (provide an appropriate reference(s)) for measuring the performances of a classifier:
  - True Positive (TP) rate,

    - False Positive (FP) rate,
    - Precision,
    - Recall,
    - F-Measure,
    - Receiver Operating Characteristic (ROC) Area.
    - Confusion Matrix.
- When you run for 5 FCV and 10 FCV, for each describe the performance of the classifiers in terms of
    - Correctly Classified Instances,
    - For each output class, report the following:
        - TP Rate,
        - FP Rate,
        - Precision,
        - Recall,
        - F-Measure,
        - ROC Area
    - Confusion matrix

---- x ----