Programming Assignment 1

Due Date: 6th February 2018, Tuesday 11.59 pm IST

---

1. Task 1
    a. Download SPECT Heart Dataset from https://archive.ics.uci.edu/ml/datasets/spect+heart
    b. Understand the dataset (i.e., number of samples, number of classes, feature values, type of features (nominal/categorical?), etc.
    c. Load the dataset into the Matlab/Python and print first instances of the dataset.
    d. Verify the number of records, the dimension of features are the same as provided in the website.
    e. Divide the total available data into three sets: Training, validation and test set. This should be run at random such that 50% is contained in training set, 25% in validation set and 25% is in test set.
    f. What do you mean by cross validation? Write a short note on cross validation. Also learn about the cvpartition function in Matlab (or equivalent in python). i.e., you should report its usage and syntax.
    g. Define the following terms (including equation): precision, recall, sensitivity, specificity, F Measure and MCC (Matthews Correlation Coefficient). Also report, the importance of each of these.
2. Task 2
    a. Just plot the features across the classes (use plot each feature separately) and pick 5 features which you feel roughly obeys Gaussian.
    b. Assume that these features selected forms a multi variate Gaussian. Thus you find the parameters characterising the distribution using MLE (Max. Likelihood Estimate: you may directly apply the equation for Mean and Covariance Matrix).
    c. Now build a Bayes classifier using only the training set and report the classification accuracy on training data.
    d. Now use the validation set, and report the accuracy. Compare the validation accuracy with the accuracy on the test set and on the training set. What inference can you make out of this comparison?
    e. For the classification system built, find the measures reported in 1.g.
    f. Now read about Naïve Bayes. Treating each features independent in determining the true class, find the 1D mean and 1D variance of each features separately.
    g. Use the Naïve Bayes rule and perform classification listed in 2.d and 2.e.

3. Submission Instructions : You should submit a zip/rar file containing 2 files (a & b)
    a. A pdf document: Each task and your answer/inference if any.
    b. A folder containing all the programs and the dataset used for the study.