CS5691: Pattern Recognition and Machine Learning
Programming Assignment 2

Deadline: February 27, 11:59 pm

You will be working in teams for this assignment. The Team details can be found here.

A) Regression:

Datasets:
Dataset 1: 1-dimensional input data, 200 samples with input range: [0.0, 5.0]
Dataset 2: 2-dimensional input data, 1000 samples with input range: [-1.0, 1.0]

Task:
> Least square regression
> Ridge regression

Languages allowed: Matlab, Python, C, C++.

Presentation of Results:
> Plot of the **approximated functions** obtained using training datasets of **different sizes**, for **different model complexities,** and for **different values of regularization parameter**. (Similar to Figures 1.4, 1.6, and 1.7 in Bishop's book).
> Scatter plot with target output on the x-axis and model output on the y-axis for the best performing model, for training and development data.
> The error values for train and development data.

Guidelines:
> Sample uniformly the subset of train and development data before starting experimentations.
> For a given sample size, use the same set of train and development data in different experiments. Experiment with different train sample sizes (e.g. N=10, 20, 50)
> Each dataset has train and development sets. Use the train set to train the model and use the development set to finetune the hyperparameters. **A blind test set will be given to each team only during the assignment vivas.** You are supposed to show results on the test data during the viva.
> A single report by a team should also include the details of the models used and observations about the results of studies. The report should be properly formatted. The sections should be numbered. The figures should be numbered and should have proper Captions.

Tip:
Read Section 1.1 of Bishop's book carefully (especially Figure 1.4 to Figure 1.8 and Table 1.2). Analyses of these kinds are expected in this assignment. You are expected to perform as many experiments as possible and provide your important observations and inferences.

B) Bayesian Classifier

You are supposed to build Bayesian classifiers (modeling each class by unimodal Gaussian density functions) for the datasets assigned to your group. This assignment is more focused on the analysis of classification techniques and getting used to handling data in Machine Learning instead of getting classification accuracy (as you will be getting pretty good accuracy in most of the cases).

Builds Bayesian models using the given dataset for the given case numbers.

Case: 1 -- Bayes with Covariance same for all classes
2 -- Bayes with Covariance different for all classes
3 -- Naive Bayes with C = \sigma^2*I.
4 -- Naive Bayes with C same for all classes.
5 -- Naive Bayes with C different for all classes.

Dataset:

Dataset for each group can be found here. The sample plots required for the assignment can be found here. Three different datasets have been provided for each team as follows:

1. Linearly separable data
2. Non Linearly separable data
3. Real Data

The first two values in each row make a 2D feature vector and the last value in each row is the class label.

Plots Required (refer to "Sample Plots" folder in the above link):

1. The plot of PDF (Gaussians) for each of your classes.
2. Decision boundary and decision surface.
3. Constant Density Curves and EigenVectors. (can be shown in plot1 or in plot2. No need for a separate plot. Refer to images 1 and 2 in "Sample Plots" for details)
4. Confusion matrices. (you may not show all in the report)
5. ROC and DET curves (they help you to compare your models). Plot them in a comparative way i.e the ROC for the models that you want to compare should be on the same plot. Similarly with DET. (Again, you may not show all of them in your report)

Libraries Allowed:

You are not allowed to use any inbuilt function for the following cases.

- You cannot use any inbuilt function for building the model or classification using the model.
- We expect you to implement the 2D Gaussian PDF that you will be using in your Bayesian classifier.

- You are supposed to write a function for plotting the ROC curve.
- Apart from this, you are allowed to use the inbuilt functions/libraries for plotting and other purposes.
- You can use any inbuilt library for plotting the DET curves.