Odd Semester 2019-2020
Programming Assignment 1
CS401-Introduction to Machine Learning

Date: 24/10/2019

Note: Each group should take the data assigned to the group only

Regression Tasks:

Dataset 1:

(a) Univariate data
(b) Bivariate data

Note: There are many files in each respective folder. Consider only the files named groupx_train, groupx_val and groupx_test as the training data, validation data and test data. In each file the last column corresponds to the actual function value

Models to be built:

1. Linear regression for Dataset I (a)
2. Multiple linear regression for Dataset I (b)
3. Polynomial curve fitting for Dataset I (a)
4. Polynomial regression for Dataset I (b)

Results:

(1) Plot of training error and validation error whenever there is a need for tuning a model parameter
(2) Scatter plot with target output on x-axis and model output on y-axis, for training data, validation data and test data
(3) Plot of the desired function and the actual function for the training data, validation data and test data
(4) Prediction accuracy of the regression model using root mean squared error

Classification Tasks:

Dataset 2: 2-dimensional data of 3 or 4 classes:

(a) Linearly separable data set
(b) Nonlinearly separable data set
(c) Overlapping data set

Models to be built:

1. Bayes classifier
   a. Covariance matrix for all the classes is the same and is $\sigma 2I$
      You can obtain the same Covariance matrix for all the classes by taking the average of Covariance matrices of all the classes. You can obtain same variance by averaging all the variances.
   b. Full Covariance matrix for all the classes and is same for all the classes
      i. You can obtain the same Covariance matrix for all the classes by taking average of Covariance matrices of all the classes
      ii. You can obtain the same covariance matrix for all the classes by taking training data of all the classes combined.
   c. Covariance matrix is diagonal and is different for each class
   d. Full covariance matrix for each class is different
2. Perceptron based classifier for Dataset 2(a)
3. Logistic regression based classifier for Dataset 2(a), 2(b) and 2(c)
4. SVM based classifier for Dataset 2(a), 2(b) and 2(c)

Report should include the results of studies presented in the following forms for each classifier and for each dataset:

1. Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data
2. Confusion matrix based on the performance for test data
3. Decision region plot for every pair of classes with the respective training data superimposed
4. Decision region plot for all the classes together with the training data superposed

Report should also include your observations about the performance and the nature of decision surface for each classifier, and for each dataset.

The report should be submitted in PDF form only. Submit your code and report strictly as one zip file by sending a mail to veenat@nitgoa.ac.in. Name the zip file as Group<num>_Assignment1_IML.zip. Eg. Group01_Assignment1_IML.zip

Deadline for submission of report: 11:55 PM, Sunday 10th November 2019