National Institute of Technology Calicut

Department of Computer Science and Engineering

CS6302E Theoretical Foundations of Machine Learning - Winter 2024

Assignment - Max. Marks: 20

Instructions: (1) Create a PDF file containing answers/results/plots for each question and give proper justification wherever necessary. Upload a ZIP file containing your source codes, and the above PDF file.
(2) Due date for uploading: 11PM, 09 April 2024

Q1

Q1.a. Generate a bi-variate normal distribution assuming a two dimensional random vector. Accept the values of parameters required to generate it from the user. Plot the distribution obtained also.
Q1.b. Create a random sample set (S1) of 800 size from this distribution and plot it.
Q1.c. For the sample obtained, what would be size of the required hypothesis class for a maximum error of 0.0, 0.05, 0.1, 0.3, 0.5, 0.8 and 1.0.

Q2

Q2.a. Divide the sample set into a train set and test set using the 80:20 method.
Q2.b. Generate a linear regression model using the train set, plot the model obtained and report the error in prediction for both the train and test set. Give the parametric form of the model also.
Q2.c. Randomly select 700 points from the 800 sample set obtained in **Q1.b** and assign "C1" as their class label and let "C2" be the class label for the remaining 100 points. Plot C1 and C2 classes. Combine these points into a new sample set S2 of size 800.
Q2.d. Create a train set of size 600 by randomly sampling S2 and let the remaining 200 samples make the test set. Make a multi-layer perceptron model to classify the C1 and C2 points. Report the confusion matrix, accuracy, precision, recall and F1-score. Plot the ROC curve and report the AUC value also. Give the parametric form of the model obtained also.
Q2.e. Carry out a k-fold stratified cross validation and report the various performance metric values for the same number of train and test set samples. Compare these values with that obtained in **Q2.d**.

Q3

Q3.a. For the above model, plot the loss function against any one of the weights from input-to-hidden edges and any one of the weights from hidden-to-output edges. Apply 5 different learning rates and compare the loss plots obtained.
Q3.b. Repeat **Q2.d.** for two more different random initialization of weights and compare the results obtained with that obtained in **Q2.d.**.
