CS:5007 Deep Learning

Feb - May 2021

Programming Assignment - 3

Classification using Perceptron

Due Date: 14th Mar 2021 11:59 PM

Follow the instructions given below carefully:

1. You are allowed to use ONLY those python libraries (libraries means NumPy etc.) which were used in today's demo.

2. You must submit your code in a **single python .ipynb notebook** with naming format as follows: Firstname_Lastname_assignment3.ipynb

3. Your code must be properly commented explaining each step clearly.

4. For each question, you must follow the instructions mentioned in it.

5. Penalty will be there if any of the above instructions are not followed.

6. Zero marks will be provided for plagiarised code.

7. NOTE: This assignment is GRADED and the total points for this assignment is 5.

Questions

1. (5 points) Implement Perceptron model from scratch and compare its performance with Logistic Regression and Naive Bayes Classifier for the task of binary classification on the given three datasets. You are **NOT** allowed to use in-built function for implementing Perceptron. For each of the three datasets, You are given two separate files, for example, data1_train.csv and data1_test.csv, having comma separated values. For data1_train.csv, you are given the features and the label for each datapoint. For data1_test.csv, you are only given the features. You will train your model on the training dataset and use the trained Perceptron model to predict labels for the datapoints in the test data given in data1_test.csv. You have to use cross validation (CV) to find best generalization accuracy. You are free to use in-built CV function from sklearn. The evaluation will be on following points:

   (a) (1 point) Correct (error free) code that we are able to run.

   (b) (1 point) Submit the predicted labels for all three datasets as follows:
   For example, for data1, the predicted labels must be put as a separate column titled 'Predictions' beside the test datapoints in data1_test.csv. Submit the modified .csv file containing the predictions. Do this for all three datasets, i.e., your submission must include **three .csv files** data_test.csv, data2_test.csv and data3_test.csv.

   (c) (3 points) Write a report such that:
   For each dataset, include the following in the report:

      i. Comparison between the performance (use metric 'test accuracy') of Logistic Regression, Naive Bayes Classifier and Perceptron.

      ii. Plot the decision boundary plots for all the three models, i.e., Logistic Regression, Naive Bayes Classifier and Perceptron, and write down your observations comparing and contrasting between the three plots. **NOTE**:You need NOT do this for dataset 3 as it contains more than 3 features.

      iii. Mention whether the data is linearly separable or not linearly separable using the Perceptron model. **NOTE**: You must submit the report in **PDF format** with the name of the file as 'Report.pdf'.