PRNN 2023 - Assignment3

Prathosh A. P.

April 6, 2023

- DataSets: All the Classification datasets provided in Assignment 1.
- Metrics to be Computed for Test Data: Accuracy, Precision, Recall, F1-Score, Confusion Matrix, Loss curves and other plots as applicable.

Problems:

1. Implement classification trees using Gini impurity and cross-entropy as impurity functions with different depths. For the MNIST problem, consider the PCA data.
2. Implement Random forest Algorithm with varying numbers of trees and features and report your observations.
3. Implement the Adaboost algorithm with at least 3 learners and one of them must be a Neural Network (MLP/CNN). Report the comparison between this and using only one classifier. Plot the convergence of train error as a function of the number of learners.
4. Consider the KMNIST data and implement (a) GMM-based clustering, (b) K means clustering. Evaluate and compare the Normalized Mutual Information for both algorithms. Experiment with different number of cluster sizes and plot the t-sne plots for all cases.
5. Implement Principal Component Analysis on KMNIST. Plot the data variance as a function of the number of principal components.

General Instructions:

1. All the data files can be found here - data
2. No ML library other than **numpy** and **matplotlib** should be used, failing which will attract zero marks.
3. A 4-6 page report has to be submitted that would list all the experiments, results, and your observations. It should be in double-column format in latex as specified here template. IISc has a subscription to overleaf and the report should be in the exact same format.

4. Use matplotlib for plotting the loss and RoC curves.

5. The final evaluation **does not** depend on the accuracy metrics but is based on the **quality of your experiments and observations thereof**.

6. We will run a plagiarism check on both your report and the codes. Any suspicion of copying would lead to a harsh penalty from negative marks in the assignment to a failing grade in the course, depending upon the severity. Therefore, kindly refrain from copying others' codes and/or reports.