SML Assignment-3

Brief Description

Imported the requisite libraries. Loaded mnist.npz. Preprocessed the dataset like last time. Selected classes 0,1 and 2. Applied PCA. Created a Decision Tree class, created functions for gini, splitting, fitting and predicting, to determine the best split acc. to the lowest gini, and fit it to the training set. Used it for prediction on x_test and found total and class wise accuracy. Implemented mostcommon function for later use in bagging (since .count() isn't allowed for ndarrays) implemented bagging with ~ 60% sampled with replacement, created 5 new trees and implement multiclass voting to get the bagging predictions. And repeated much of the latter part of the second subquestion for the accuracies.

These results were operated with a seed value of 16 for np.random

Overall accuracy is - 0.8630441690498888
Classwise accuracy is- [0.9561224489795919, 0.7629955947136564, 0.8846899224806202]
Bagging overall- 0.9408960915157293
Bagging classwise- [0.9622448979591837, 0.9550660792951542, 0.9050387596899225]

Process finished with exit code 0