



In the name of God
Machine Learning Course (Spring 2021)
Assignment #3: Ensemble Algorithms

Due date: Ordibehesht 16th

In this assignment you are to implement two ensemble algorithms, Bagging and AdaBoost.M1. This homework should be implemented in Python. So, you need to install Scikitlearn and Numpy packages if you need to.

As you know these algorithms need base classifiers, for example decision trees. You need to use `sklearn.tree.DecisionTreeClassifier(...)` classifier of Scikitlearn package as the base learner of you ensemble models.

Note1: `max_depth` parameter of the base learner in AdaBoost.M1 algorithm should be tuned **experimentally** so that the decision tree performs a little better than a random classifier. For the Bagging algorithm, use the default parameters of the base learner decision tree.

Add Gaussian noise with the following parameters $\mathcal{N} \sim (0, 1)$ to 10%, 20%, and 30% of the features randomly on each data set and compare the results with noiseless setting.

You should split each data set to train and test parts. Use 70% of the data for training phase and the remaining 30% for testing phase. Run your codes for 10 individual runs and report the average **test accuracies** of 10 runs on each data set.

Note2: the iteration number T in Bagging and AdaBoost.M1 algorithms should be obtained from sets $\{11, 21, 31, 41\}$ and $\{21, 31, 41, 51\}$, respectively. In other words, you should test the performance of the algorithms with the given T values of each algorithm and report your best results over a fixed T value.

Please use the attached book for the pseudo code of the algorithms.

Use these following data sets in your homework and report your results in the following tables. You should provide these tables in your report and fill them with your results in each part.

Datasets	Algorithms	
	Bagging	AdaBoost.M1
Wine		
Glass		
BreastTissue		
Diabetes		
Sonar		
Ionosphere		

Datasets	Algorithms	
	Bagging + noise 10%, 20%, 30%	AdaBoost.M1 + noise 10%, 20%, 30%
Wine		
Glass		
BreastTissue		
Diabetes		
Sonar		
Ionosphere		

Questions: (Answer these questions in your report)

- Why should we set max_depth parameter in AdaBoost.M1 so that the base classifiers become a little better than random?
- What do we mean by stable, unstable, and weak classifier?
- What kind of classifiers should be used in Bagging? How about AdaBoost.M1?
- Compare the results in noiseless and noisy settings and say which algorithm's performance degrades with noisy features. And why?

Important Notes:

- You need to implement the algorithms from scratch. Using the built-in functions and algorithms is not allowed.
- Feel free to use your preferred programming languages.
- Pay extra attention to the due date. It will not be extended.
- Be advised that submissions after the deadline would not grade.
- Provide a report for your assignment and explain your features, code, and results.
- The name of the uploading file should be your **Lastname_Firstname**.
- Using other students' codes or the codes available on the internet will lead to zero grades.

Good Luck