MSDS 630
HW 4
AdaBoost
Due: Feb 24, 2025 @ 11:59p

1 Exploring AdaBoost (10 pts)

Now we will apply AdaBoost to classify a toy dataset. The dataset consists of 4 points: $(x_1 = (0, -1), y_1 = -1), (x_2 = (1, 0), y_2 = 1), (x_3 = (-1, 0), y_3 = 1)$and $(x_4 = (0, 1), y_4) = -1)$. You may want to use Python as a calculator rather than doing the computations by hand but you don't have to submit your code.

1. (6 points) For M = 4 (use 4 trees), show how Adaboost works for this dataset, using simple decision stumps as weak classifiers. For each step fill the following table:

m	w_1	w_2	w_3	w_4	err	α	$T_m(x_1)$	$T_m(x_2)$	$T_m(x_3)$	$T_m(x_4)$
1										
2										
3										
4										

2. (2 points) What is the training error of AdaBoost for this toy dataset? Show the computation.

3. (2 points) Is the above dataset linearly separable? Explain why AdaBoost does better than a decision stump on the above dataset.

   Note: In the AdaBoost algorithm all log functions are natural log and not log 10.

2 Implementing AdaBoost (15 pts)

2.1 Instructions:

= For this exercise you will implement AdaBoost from scratch and applied it to a spam dataset. You will be classifying data into spam and not spam. You can call DecisionTreeClassifier from sklearn (with default maxdepth=1) to learn your base classifiers. Write a program implementing AdaBoost with trees using the template and tests given to you adaboost.py. Here is how you train a decision tree classifier with weights.

```
h = DecisionTreeClassifier(max_depth=1, random_state=0)
h.fit(X, Y, sample_weight=w)
```

- Starting code can be found in `adaboost.py`. Some functions are pre-written according to their descriptions, others need to be written by you. These functions are indicated with sections like:

```
### BEGIN SOLUTION
```

```
### END SOLUTION
```

The specific directions for each function are found in the function docstring.

- Do not submit full data to Github. Doing so will result in a loss of 5 points. Only commit the tiny datafiles.

2.2 Deliverables:

Submit your completed functions as a Python script entitled "`adaboost.py`".

2.3 Evaluation:

Once you've finished, make sure you can run:

```
pytest test_adaboost.py
```

There are 3 tests to pass. Each pass is worth 5 points. If you fail a test, but your code is on the right track, half credit will be awarded for that test.