3684 – Advanced Topics in Machine Learning, Spring 2022
Home Assignment #1a – Boosting demonstration

Lecturer: Dr. Leon Anavy
Teaching Assistant: Mr. Alon Oring

General instructions:

1. Submission is **individual.**
2. Submission must include python code **and** a written report.
3. You may use external libraries. Specify all required libraries in a proper manner.
4. Your code must be reproducible. Code that will not run will result in a grade reduction.
5. Your report should be clear, coherent, and concise. The report should not exceed 10 pages.
6. Invest thoughts and considerations to the way you choose to present data and experimental results.
7. All figure and plots should include captions, labels and data units. Pay attention to data visualization guidelines.

Assignment tasks:

The goal of this assignment is to demonstrate two concepts that were covered in class.

1. Boosting (General)
   a. Generate a dataset with two features and one binary label.
   b. Present the dataset on a 3D space (in the label space) in a similar way to the demonstration we did in class
   c. Demonstrate the Boosting algorithm by:
      a. Implement a weak classifier (A random guess with error < 0.5 is a good option)
      b. Iteratively add the classifiers to the ensemble with a small constant α.
      c. Present the result of the ensemble on the 3D space.
      d. Present the loss and the training error of the ensemble of each iteration
   d. Experiment with different number of samples (choose 3 samples for the visualization)
2. AdaBoost
   a. Generate a dataset with two features and one binary label. Either randomly, manually, or with some underlying model.
   b. Show the sample on the 2D plane of feature space.
   c. Implement AdaBoost by:
      a. Implement the selection of a decision stump
      b. Implement the calculation of the sample weights
      c. Implement the calculation of the step size
   d. Run AdaBoost and
      a. Present the weight of each sample
      b. Present the current decision boundary and also the "certainty" of each region (the weighted average of the decision stumps).
      c. Present the loss and the training error
      d. Present a test error
   e. Experiment with different number of samples (choose 3 samples for the visualization) and different datasets

Summarize all your work in a scientific/professional report.

Class presentation:

If you have chosen this assignment as your class presentation assignment you are required to prepare a 30 minutes presentation in which you will need to showcase your work. You should cover all aspects of your work in the presentation.