3945 – Advanced Machine Learning, Spring 2022
Home Assignment 2 – Ensemble Learning

Lecturers: Dr. Leon Anavy, Mr. Alon Oring

In this assignment, you will implement Ensemble Learning methods - Gradient Boosting Regression Trees (GBRT) and AdaBoost. You will demonstrate the models and design experiments to test their performance.
You will submit a single ipython notebook (.ipynb) file that includes all the code and outputs, as well as a brief report that explains your approach and findings. You are encouraged to use methods and techniques covered in class but are also welcome to use any other relevant method you find suitable for the task.

Part 0: Generate datasets

1) Generate at least two datasets with one binary label.
2) Split the data to train and test sets
3) Visualize the data (You can use dimensionality reduction if needed)

Part 1: Gradient Boosting Regression Trees (GBRT)

4) Implement GBRT
   a) Loss function
   b) Residuals calculation
   c) Weak regression tree
   d) Boosting step
5) Demonstrate your GBRT model on the datasets you generated
6) Design and perform experiments to test the performance of your model. Present and discuss the results of the experiments.

Part 2: AdaBoost

7) Implement AdaBoost
   a) Weights (w_i)
   b) Decision stump
   c) Weighted classification error (ϵ)
   d) Step size (α)
   e) Boosting step
   f) Updating weights
8) Demonstrate your AdaBoost model on the datasets you generated
9) Design and perform experiments to test the performance of your model. Present and discuss the results of the experiments.

Submission Guidelines

- Submit the work in pairs. Only one submission for each pair.
- Your submission should include a single zip file containing:
  - A single ipython notebook (.ipynb) file that includes all the code and outputs.
  - A brief report (1-2 pages) in a pdf format that explains your approach and findings.
- The submitted file should follow the naming convention:
  - 3945_HW##_XXX_YYY.zip
    Where:
    - ## is the assignment number
    - XXX and YYY are your student numbers (IDs)

    For example: 3945_HW1_123456789_987654321.zip
- Make sure to run your notebook from start to finish before submitting to ensure that it runs without errors.
- You may use external libraries. Specify all required libraries in a proper manner.
- Grading will be based on correctness, elegance of solution, and style (comments, naming conventions, etc.)
- Your report should be clear, coherent, and concise.
- All figure and plots should include captions, labels and data units. Pay attention to data visualization guidelines.
- Make sure to use correct ML methodologies and justify your selections (split the data to train/test, tune hyperparameters, report relevant performance measures).