Report

The following experimental observations were conducted on my local Machine running MacOS - Catalina.

Specifications: 2.3 GHz Dual Core Intel Core i5 Processor with 8 GB RAM.

Perceptron

For the linearly-separable dataset:

1. Average runtime of the perceptron algorithm under ERM is around 0.065 seconds.
2. Average runtime of the perceptron algorithm under 10 fold cross-validation is around 0.65 seconds.
3. The algorithm terminates and reaches convergence in around 120 iterations.
4. Perceptron achieves 0% ERM error on this dataset and 0.2% mean cross validation error.

**For the breast-cancer dataset:**

1. The algorithm doesn't terminate by itself on this dataset since the data is not linearly separable and vanilla perceptron works only on linearly separable datasets.
2. A hard bound threshold of 1000 iterations of perceptron is used as a heuristic to terminate in case the data is not linearly separable.
3. Average runtime of the perceptron algorithm under ERM is around 0.25 seconds.
4. Average runtime of the perceptron algorithm under 10 fold cross-validation is around 2.5 seconds.
5. Perceptron achieves 22.8% ERM error and 22.6% mean cross validation error.

Adaboost

ERM

The diagram on the right shows the plot of training error while performing ERM with different number of rounds of Adaboost. Clearly ERM error with adaboost converges to a minimum after around 96 iterations.

I recommend around 100 iterations of Adaboost for performing ERM.

Plot of Empirical Risk with T rounds of Adaboost

Plot of Mean Validation Error with T rounds of Adaboost
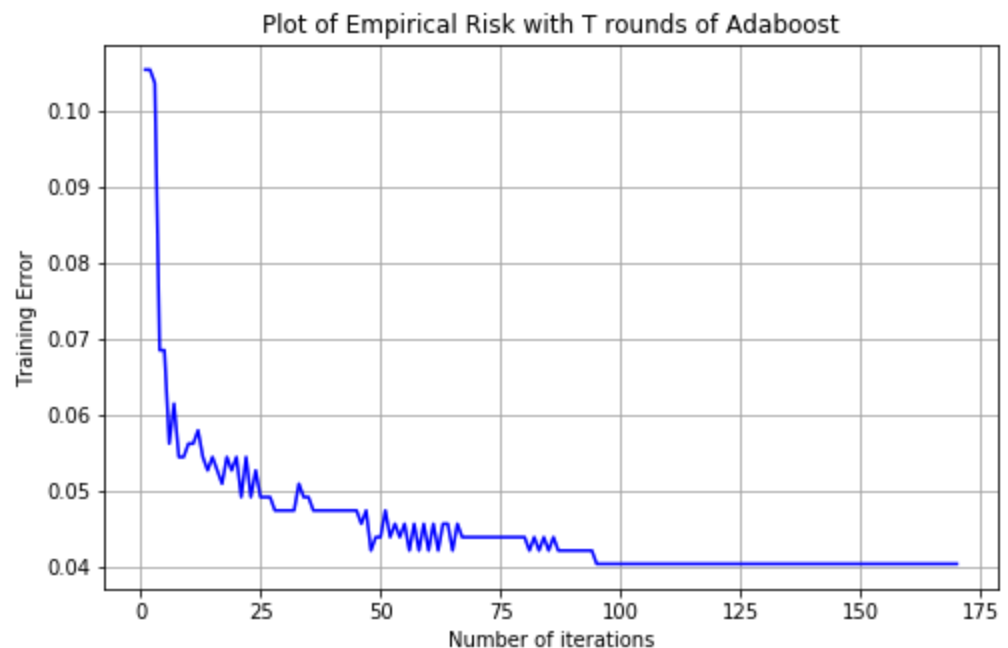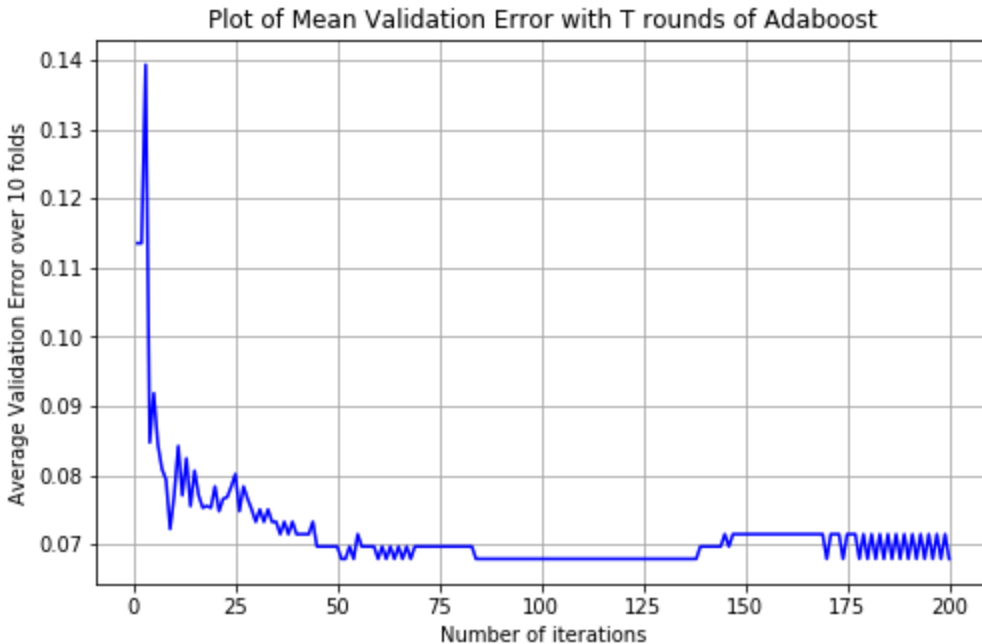
Cross-Validation

The above diagram shows the plot of mean cross-validation error across 10 folds versus number of iterations of Adaboost.

Clearly the error starts to oscillate as the number of iterations increase but it never reaches a global minimum. For this task I would recommend 85 iterations of Adaboost. This achieves an average accuracy of 93.3% across 10 folds.

README

The instructions below will get you started with running the programming component of the Assignment on your local machine. The below commands work for Linux/MacOS distributions.

Prerequisites - Python3

Open a Terminal/Shell and run the following commands. The dataset input to the given command should be a CSV file.

perceptron

Command : python perceptron.py --dataset <path to dataset.csv> --mode <erm/cv>

Output : Weight vectors from the perceptron algorithm and mean error.

adaboost

Command : python adaboost.py --dataset <path to dataset.csv> --mode <erm/cv>

Output : Weight vectors from the adaboost algorithm and mean error.