Assignment

Implement an Adaboost classifier. For each week, your feature set is (μ, σ) for that week. Use your labels (you will have 52 labels per year for each week) from year 1 to train your classifier and predict labels for year 2. Recall that are two hyper-parameters in the random forest classifier

1. N - number of "weak" learners to use
2. d - base learner (base estimator)
3. learning rate λ

Questions:

1. take $\lambda = 0.5$ and $\lambda = 1$. For each lambda, construct an Adaboost classifier with any three base estimators of your choice (e.g. logistic regression, naive bayesian, k-NN). Use your year 1 labels as training set and compute the error rate for year 2. Plot your error rates as you change N from 1 to 15.
2. for each base estimator, what is the best value N^* for learning rate $\lambda = 0.5$?
3. what is your accuracy for each base estimator choice (assuming the best N^* for that estimator)

4. what classifier is best to use as base estimator for your data?

5. implement a trading strategy (using the Adaboost with the best estimator) based on your labels for year 2 and compare the performance with the "buy-and-hold" strategy. Which strategy results in a larger amount at the end of the year?