Assignment

Implement a random forest 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 (sub)trees to use
2. d - max depth of each subtree

Questions:

1. take $N = 1, \ldots, 10$ and $d = 1, 2, \ldots, 5$. For each value of N and d construct a random tree classifier (use "entropy" as splitting criteria - this is the default) use your year 1 labels as training set and compute the error rate for year 2. Plot your error rates and find the best combination of N and d.
2. using the optimal values from year 1, compute the confusion matrix for year 2
3. what is true positive rate and true negative rate for year 2?

4. implement a trading strategy 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?