The basic idea is to separate modeling into two parts and put them back together when they’re done.

Part 1: Getting a specific weight on each sector by feeding historical weights of the most successful hedge funds (Dataset can be achieved from Whalewisdom)

Part 2: Getting a yearly or quarterly return for each index ETF that will represent each sector in Part 1.

At the end of the day, we can combine the results from each part into one to get an idea of what weights make the optimized sector allocation for our individual portfolio.

Structures

Part 1:

1. Determine which funds to be selected for our project

(Could be top-n (3, 5, 10..) best performed ones or other standards that we may consider proper)

1. Pulling the quarterly 13-F data from Whalewisdom for the selected funds
2. Define the features and outputs.

(Ideally features should be weightings on each sector quarterly that spitting out optimized weightings for output that would have given the best portfolio return over that time periods. However, building this kind of model could be difficult to implement.)

Part 2:

1. Define which index ETF matching with which sector in Part 1.
2. Determine the period that we are going to analyze (probably starting with OCT 1, 2018 when the reclassification started)
3. Pulling each ETF’s historical data from APIs (Alpaca)
4. Data preprocessing with the data. Need to apply the time-series way of splitting it.

(Features = daily closing price, Output = daily closing price (using rolling window))

1. Build and Train the model using RNN LSTM model
2. Predict the data using test dataset
3. Evaluate the model

After completing both parts, we can simulate a portfolio by using the weights recommended from Part 1, applying them to relevant ETFs to see the result (prediction).

Since we don’t know the exact dates when the hedge funds rebalanced their weights on each sector during a certain quarter, we need to assume that they did in the middle of the quarter(45 days after the quarter starts) for our simulated portfolio.

When we get a predicted result of return from our simulation portfolio, then we should compare the result to that of the real funds to see how relevant the result is. If it is relevant enough, we might use this method to create our real-world portfolio.