Stats 551 Project Proposal

Forecasting Daily Stock Prices of Apple with Bayesian modeling

Elishua Shumpert, Josh Jiang, and MyungJin Lee

The stock market is historically known to go through phases of generally rising prices and lowering prices. While these trends can be easily identifiable in hindsight, what state the stock market is in currently and where it will go from here is a problem of great interest to investors since the inception of the stock market. From a Bayesian perspective, the stock market can be seen as existing in one of two un-observable states, one which is in a general decline (bear market) and another that is in a general uptrend (bull market), while the movement of the market is a random realization of the underlying state. In effect, we described the stock market as a hidden Markov model (the number of hidden states can be optimized, we used two here as an example).

In this research project, we plan to utilize a hidden Markov model to identify the hidden state of the stock market. By identifying the state of the market, we can get a general idea of where market is going and that can be used to influence investing strategies. The S&P 500 is generally considered to be a barometer for the health of the entire stock market so we will use SPY, an ETF that tracks that S&P 500 for our analysis. We will using daily open, high, low, close, and volume historical data as input for our model as well as for back-testing. When fitting the model, we plan to have a sliding window of historic data. For instance, if we are attempting to find the hidden state of the stock market on t date, we will will fit the model on the data from $t - \delta$ to t, where δ is some time interval we have yet to decide (somewhere in the order of years).

We will use a Gaussian HMM, which can be described mathematically as

$$\lambda = \{A, \mu, \sigma, \pi\},$$

where A is the transition matrix represents the probability of transitioning from one state to another, μ and σ represent the mean and variance of observations with respect to each hidden state, and π is the vector of probabilities of initial starting states. When it comes to training an HMM on real-world data, there exists three main sub-problems:

1. Estimate the probability of occurrence for the set of observations.
2. Determine the optimal sequence of hidden states for the HMM given the set of observations
3. Find the optimal parameters A, μ, σ, π of the HMM.

There is extensive literature and research done on each of these sub-problems. Furthermore, there are well-known packages that can implement the techniques to solve each sub-problem. At the moment, we have not done any preliminary data analysis. Our plan is to complete:

- Exploratory data analysis and creation of HMM by March 1st

- Make adjustments to model and analyze results April 8th
- The final draft and final report by April 15th
- The presentation prepared by April 19th.

Because we are only working with one model, we believe it would be most efficient for one person to code up the HMM, Josh will be in charge of that. Eli will be in charge of exploratory data analysis and data pre-processing. MyungJin will be in charge of background research and writing the final report. We will call meetings and touch base will each other as needed throughout the project.