ISyE 6416 – Basic Statistical Methods - Spring 2016
Bonus Project: “Big” Data Analytics
Proposal

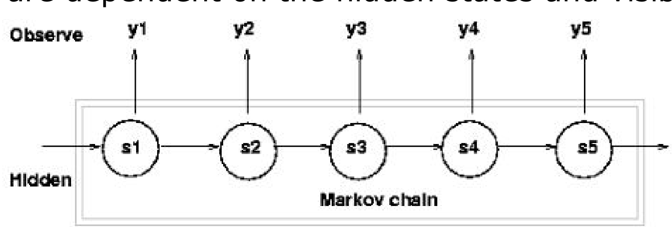
Team Member Names: Ziqi Yang, Feng Gao

Project Title: Hidden Markov Model for Stock Market Index

Problem Statement
In financial world, the market behaves differently when it is in “good” state, “bad” state, or “normal” state. One key difference is that the market volatility, which can be represented by the volatility of SP500 index, varies by states. Usually, when the market is doing badly, the volatility will be higher than it is doing better. Acknowledging which states we are in is important to all market participants such as financial institutes and investors, since it was not only a signal to what the market will behave, but the volatility predicted by state also can be used for VIX (CBOE volatility index) pricing, option pricing and so on. In this project, we are going to use Hidden Markov Model to analyze which state the market is in and trying to obtain the volatility of each states and using the Black-Sholes formula for option pricing.

Data Source
The data we will use is the SP500 index from Jan 1st, 1993 to Dec 31st, 2015, which can be obtained from Yahoo Finance. When pricing option, another input is the risk free rate. We can use the 6 months Treasury bill and it can be obtained from US. Department of Treasury.

Methodology:
1. Hidden Markov Model
In hidden Markov model, the system is assumed to be a Markov process with hidden states. The underlying states are unobserved and follow Markov chain process with certain transition probability. Observations are dependent on the hidden states and visible.



Graph 1: Hidden Markov Model

Graph 1 shows the process of hidden Markov model with S being the hidden states and y being the observations.

2. Black-Scholes Formula
In Black-Scholes model, stock prices follow geometric Brownian Motion. The price process of the underlying stock is

$$S_t = S_0 \exp[\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma B_t]$$

where S_t is the stock price at time t, S_0 is the initial stock price, μ is the drift, σ is the volatility and B_t is the Brownian Motion. Taking logarithm on both sides of the equation above we get

$$\log(S_t) = \log(S_0) + \left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t$$

The logarithm of stock price at time t follows the normal distribution with expectation $\mathbb{E}\log(S_t) = \log(S_0) + (\mu - \sigma^2/2)t$. The volatility of stock price σ and the drift μ are assumed to remain constant across time in Black-Scholes model. However, in reality, they will change with the overall conditions of the stock market. Combined with the Markov chain of economic situation, Black-Scholes model is modified to be Markov Black-Scholes model, which can accommodate the change in states of the underlying financial assets.

3. *Forward-Backward Algorithm*

The forward-backward algorithm is developed to compute the posterior marginal of the hidden state variables based on observations. Since the parameters of the Markov Black-Scholes model including the drift μ, the volatility σ and the transition probability P are unknown, we apply forward-backward algorithm, which can help to estimate these parameters. The algorithm has two passes, in which the forward probability and backward probability are calculated respectively. The forward probability is represented as

$$\alpha_k(t) = \mathbb{P}(O_1, \dots, O_t, I_t = k)$$

and backward probability is represented as

$$\beta_k(t) = \mathbb{P}(O_{t+1}, \dots, O_T, I_t = k)$$

where O_j is the observation at time j and I_j stands for the state of the underlying stock at time j. The forward probability $\alpha_k(t)$ and backward probability $\beta_k(t)$ can be calculated in a recursive way. Imposing the probabilities into likelihood functions and transition equations,

$$L_k(t) = \mathbb{P}(S_t = k|O) = \frac{\alpha_k(t)\beta_k(t)}{\mathbb{P}(O)} \propto \alpha_k(t)\beta_k(t)$$

$$H_{k,l}(t) = \mathbb{P}(S_t = k, S_{t+1} = l|O) = \frac{\alpha_k(t)a_{k,l}b_l(O_{t+1})\beta_l(t+1)}{\mathbb{P}(O)}$$

where $a_{k,l}$ is the probability of transiting from state k to state l and $b_l(O_{t+1})$ is the probability of observation O_{t+1} given state l. Then we can get the estimates of parameters in the normal distribution and the transition matrix P.

4. *Viterbi Algorithm in Hidden Markov Model*

The Viterbi algorithm is used to find the most likely sequence of the hidden states based on a series of observations in hidden Markov model. In this algorithm, the maximum likelihood estimation procedure is implemented in a recursive way, which makes it efficient to calculate the corresponding probabilities. In each step, the algorithm incorporates one more observation in the data series and the complexity is $O(\mathrm{kt})$ if the total number of states is k. The recursive process is shown as below.

$$\mathbb{P}(S_0, S_1 \dots S_T, O_1 \dots O_T) = \mathbb{P}(S_0)\mathbb{P}(S_1 \dots S_T, O_1 \dots O_T|S_0)$$
$$= \pi_{S_0}\mathbb{P}(S_1, O_1|S_0)\mathbb{P}(S_2 \dots S_T, O_2 \dots O_T|S_1, O_1, S_0) = \cdots = \pi_{S_0}\{\Pi_{i=1}^{T} a_{S_{i-1},S_i} b(O_i|S_i)\}$$

At each iteration, the formula has the similar form and the likelihood function finally becomes a product of T terms by induction.

Taking logarithm on both sides of the equation above, we can simplify it into

$$\log\mathbb{P}(S_0, S_1 \dots S_T, O_1 \dots O_T) = log\pi_{S_0} + \Sigma_{i=1}^{T}(loga_{S_{i-1},S_i} + logb(O_i|S_i)$$

By this transformation, the maximum likelihood problem is converted into the shortest path problem. We can adopt Dijkstra's algorithm or Bellman-Ford algorithm to find the shortest path, which is also the most likely sequence of the hidden states.

Expected Result

In this project, we expect to identify the hidden states of underlying financial assets based on the observed movement of prices. First, we should determine the different parameter sets of the drift μ and the volatility σ in different states by the forward-backward algorithm. For example, during the 2008 financial crisis, the volatility is expected to be large and the drift should be relatively low as it was a bad period for the financial market. Second, based on the parameters we have, we should uncover the sequence of the hidden states of the underlying stocks or indices during the research period and determine which years are the hidden turning points for the financial market.