Section 1: Week 6: Problem Solving with an Algorithm

Nate Bachmeier

TIM-8110: Programming Languages and Algorithms

June 9th, 2019

North Central University

# Machine Learning in Financial Systems

An author for Bloomberg Businessweek recently wrote an entitled “Computer Models Won’t Beat the Stock Market Any Time Soon.” Dewey makes several claims such as (1) it is a random walk that is ever changing; (2) there’s more signal than noise; (3) there is insufficient data; and (4) as proposed by the efficient market theory the edge is too small.

These claims raise two follow up questions (1) is there validity to the argument; and (2) what characteristics would need to exist in such a model. It is not the goal of this paper to prove of disprove the position only to explore its merits, given the broader state of scientific discovery.

Obviously if Dewey’s claims can be disproven the impact would be significant to the consumer of the computer model. They could choose economic gains or minor celebrity status within the financial community, perhaps even write a best-selling book.

# Claims

## Random Walk

In 1973 Malkiel coined the term in his book “A Random Walk Down Wall Street,” which proposes that (1) all price action is purely random; and (2) an efficient market exists where all knowledge is instantly available to all parties. Thus, the best algorithm would be passively investing in a well-diversified portfolio.

Assuming for a moment that it truly is a random walk, this is not a limitation. These scenarios “play a central role in graph theory and in the study of combinatorics, percolation theory, classical and quantum field theory and a myriad of other applications in physics and mathematics (Rudnick, Joseph, & Gaspari, 2004).”

Rudnick describes several algorithms for associating probabilities to clusters of different combinatorial sets. This is accomplished through “generating functions” paired with statistics and differential equations. While it is not possible to state, “in 6 months the S&P 500 will be X or Y,” it is possible to assign a likelihood to both values.

## Signal to Noise Ratio

There is already too much signal in the ticker tape, and then additional sources are often consulted. First, the assumption that multiple sources adds to the complexity of modeling is discounting the vast commoditization of big data technologies. Second, there is an assumption that noise cannot be efficiently removed.

Astronomical videos are frequently corrupted with ‘impulse noise’ which are bright or dark spots on arbitrary frames. This has led to the necessity for algorithms to clean up the feed.

The naïve solution is to implement a sliding window of N frames, and then take the average value of each pixel across the frame set. A better solution would be to average each pixel using only the frames where that pixel is not corrupt. This can be accomplished by (1) determine the intensity of each pixel; (2) determine the nearest neighbors; (3) construct a Similarity Filter Mask for each neighborhood; (4) apply the filter mask to omit any unexpected intensities (Aliakhmet, 2019).

A similar model could be applied to the financial markets, where (1) the price or volume represent the intensity; (2) each chart candle becomes a frame; and (3) multiple assets become different pixels within the frame. This may lead to a more efficient “moving average” and if not, there are many more algorithms being created in this field.

## Insufficient Data for Modeling

Financial markets have existed for hundreds of years, yet quality records are only widely available for the last 100. One approach to produce more training data is with the use Generative Adversarial Networks (GAN). GAN algorithms pair two deep learning algorithms in a loop where (a) creates fake data; and (b) predicts if it is legitimate. Both recursively train each other resulting in high quality forgeries.

Another approach is to reach across the sciences and find similar problems in other domains. For instance, dynamically forecasting conditional probabilities on time series is a hot topic in risk analysis. Li, Zhuang, and Shen proposed an algorithm which uses ‘partial auto correlation functions’ and ‘auto regressive integrated moving averages’ to reliably predict the expected number of terrorist bombings. It may be possible to apply their findings to financial markets, which also rely on correlated data streams to make predictions about direction and intensity.

## Edge is Insignificant

Dewey’s final claim is that despite the potential payout from building such an oracle, it would not be useful as the edge is too small. This is due to Louis Bachelier’s Theory of Speculation, which states that everyone knows everything and acts rational.

Since its publication in 1900, behavior economists have collected enough evidence to fully disprove that people are always rational, especially when money is involved.

Dalton adds that the efficient market theory argument is discounting the ‘collision of different time frames.’ During a crashing market a short-term account maybe best served by going short. However, a long-term account may see the demise as an opportunity to add to their position.

# General Structure of Desired Algorithm

There is enough evidence to conclude that a model could be created, and that it could likely beat the market with an acceptable level of consistently. The core algorithm would likely consist of steps (1a) acquire new quote frame; (1b) apply noise filter; (1c) compute moving averages; (2) determine market condition; (3) determine the desired state; (4) transition if a threshold has been exceeded.

## 1. Prepare Next Quote Frame

Each frame represents a snapshot of all relevant asset metrics, with an individual pixel containing the pricing and volume information. Like previously described algorithms, each pixel needs to be transformed based on the filtered moving average.

Additional investigations are required to determine if the average should be projected as a relative delta change or literal value. It might also be advantageous to compute the moving averages multiple times with different lengths to represent the views of different time frames. Many existing systems already leverage 20, 50, and 200 day moving averages to account for this.

## 2. Determine Market Condition

A market is defined as a collection of assets, such as technology or international stocks. Each market can be on a real number between -1.0 (short) to 1.0 (long). Each exchange publishes aggregate metrics, called market internals, which can be fed into a transform to derive this value. If an internal is not available for a basket of assets, it can be calculated with basic arithmetic.

## 3. Determine Desired State

There are five dimensions that lead to profitability (alpha) in financial markets. These are represented as the tuple (Beta, Delta, Vega, Gamma, Theta). Beta is the deviation of correlated assets; Delta is gain from 1$ increase in direction; Vega is 1% increase in volatility; Gamma is increase rate of delta per 1$ increase; and Theta is cost (interest) of 1 day.

The market condition of the desired time frame determines the optimal configuration of the tuple. During an uptrend money can be made from aligning with the trend (Delta), reducing size as it reverts downtrend (Gamma), an understanding that a positive move up decreases volatility (Vega), and reducing interest paid on the position (Theta). This desired state would be described by the tuple (0, +delta, -vega, +gamma, 0).

## 4. Transition to Desired State

If the current state S exceeds the threshold of desired state D; then a rebalance strategy is required. The rebalancing occurs through a multi-dimensional version of the Knapsack Problem, which chooses the maximum value from the fewest items.

The available items are listed below with each dimension a real number between -1.00 and 1.00. For instance, to gain 60 deltas 6 x 10 delta calls or 1 x 60 delta call are equivalent. During the minimization of the other dimensions more constraints will cause the algorithm to deside which is more desirable.

* Long Call (+delta, +vega, +gamma, -theta)
* Long Put (-delta, +vega, +gamma, -theta)
* Long Stock (+delta, 0, 0, +dividend - margin)