Machine Learning Capstone Proposal- Udacity

Hitesh Gulati
hg0111@gmail.com

Project Overview.

Basics of Investment

Investment and trading have been around us for long time even before the technology age. The two-elementary metrics to evaluate a trading decision are return and risk. Return is the extra amount received in future per cent of amount invested in present. Risk on other hand is the chance that an investment's actual return will differ from the expected return. Risk also includes the possibility of losing some or all of the original investment (investopedia.com, 2017). Most common measure o risk is standard deviation of periodic returns, higher the deviation higher is the risk. Return the extra money made is a desirable product while risk the probability of loss or uncertainty should be reduced to as little as possible. Both risk and return goes hand in hand, to generate extra return the investor must be willing to take additional risk and thus there is a trade-off between risk and return. Hence the objective of investments firms is to generate maximum return while taking on minimum risk.

Styles of Investment

Broadly there are two types of investment analysis employed by practitioners, fundamental analysis, and technical analysis. Fundamental analysis involves the estimation of intrinsic value of company's share using it's data, such as earnings and sales forecasts, and risk estimates as well as industry and economic data, such as economic growth, inflation, and interest rates. Buy and sell decisions depend on whether the current market price is less than or greater than the estimated intrinsic value (W. Sean Cleary, 2014). Technical analysis on other hand involves the analysis of historical trading information (primarily pricing and volume data) in an attempt to identify recurring patterns in the trading data that can be used to guide investment decisions (W. Sean Cleary, 2014).

Project Description

In this project I will try to create an automated trader which will use machine learning algorithm to make a decision to buy a share if it predicts increase in price, sell/short if it predicts price decrease or hold in case of risk prediction. The risk level assumed by the learner is pre-set. My personal motivation for this project is I want to make a career in financial advisory and to be able to apply machine learning concepts in this domain was the major reason I took the course. This particular idea of using multiple agent learner is from (Larsen, 2010).

Problem Statement

The industry has a rich knowledge base and many techniques have been used in past to indicate future prices. These indicators (candlesticks, moving average, average directional index and so on) usually take in historical data for stocks (Open, High, Low, Close, Volume and Adjusted Close) and outputs their signal Buy, Hold or Sell.

These indicators may have been powerful earlier but with advent of time and their overuse are no longer capable enough to predict a buy or sell single handed. The objective of this project will be to apply these various indicators (or agents) and based on their outputs, the learner decides whether to buy, sell or hold on a trade. The objective of the project will be to discover whether the learner can make better decision than all individual indicators.

These indicators can be based on fundamental or technical analysis or combination of both. However, in this project I have included only technical indicators as the data required by

fundamental analysis is not available easily. The learner takes input from these indicators and thus new indicators can be added later without disrupting the structure.

Datasets and Inputs.

The input dataset for this project will be daily trading data (Open, High, Low, Close, Adjusted Close and Volume) from Nasdaq of retail industry from year 2006 to 2013. To keep the amount of data in check I have limited myself to a single sector and that is Retail Sector. The data has been fetched from Yahoo Finance using my python script with help of pandas_datareader package (https://pandas-datareader.readthedocs.io/en/latest/).

Various agents will then be employed which take this data as input and generates a signal based on its capacity. A generic form of these agents will be:

$$\text{Agent } (P_s, t) \rightarrow \{s_1, s_2, s_3, \ldots s_n\}$$

P_s: Price series (Open, High, Low, Close, Adjusted Close, Volume data)
t: Times series data for each trading day.
$s_1, s_2, s_3, \ldots s_n$: Signals generated by the agents as output.

The agents are described:

Hodrick Prescott Trend Agent

Trend agent is used to smooth out short term fluctuations from the data and estimate a general movement trend of the series by using long term averages. Hodrick-Prescott (HP) FIlter is one such data-smoothing technique that is commonly applied to remove short-term, thereby revealing long-term trends (investopedia.com, 2017).

$$\text{Hodrick Prescit Agent } (P_s, t) \rightarrow \{\text{Uptrend, Downtrend, Notrend}\}$$

Moving Average Crossover Agent

This agent generates two moving averages one for short period (14 days default) and other for long period (35 days by default). The longer moving average will be a smoothed-out form of short moving average and will replicate its movements in case of big shift. Thus, the agent releases a signal of Buy whenever shorter average rises above longer average, a Sell signal when it drops below longer average and Hold otherwise.

$$\text{MAC Agent } (P_s, t) \rightarrow \{\text{Buy, Sell, Hold}\}$$

Candlestick Agent

Candlestick is a charting technique of time series data, Candlestick analysis is an activity of pattern recognition of candlestick charts and assign a label to the same as Buy, Sell and Hold. Our agent will try to identify some of the patterns and generate a signal of Buy, Sell, or Hold accordingly.

$$\text{Candlestick Agent } (P_s, t) \rightarrow \{\text{Buy, Sell, Hold}\}$$

Stochastic Agent

Another type of agents study the momentum of the trend to identify when a stock is overbought or oversold. An overbought stock is a stock whose price has increased drastically over a short period of time and is trading at an artificially high price compared to recent price activity. If a stock is trading at an artificially high price the theory suggests that the price may reverse, thereby generating a sell signal (Larsen, 2010). Opposite for oversold stock.

$$\text{Stochastic Agent } (P_s, t) \rightarrow \{\text{Buy, Sell, Hold}\}$$

Volume Agent

Volume means total number of shares traded of a stock in a given time frame and this can be an important source of information. Primary objective of this agent is to measure current volume compared to previous volume and return a signal as strong volume or weak volume.

$$\text{Volume Agent } (P_s, t) \rightarrow \{\text{Strong Volume, Weak Volume}\}$$

Average Directional Index

Average directional index is used to measure the strength of the trend and not necessarily the actual direction of it. This index can be used to determine if a security is trending or not. This agent is added in the project as it could add synergy when used with other agent like trend agents.

$$\text{Average Direction Index Agent } (P_s, t) \rightarrow \{\text{Strong Trend, Weak Trend}\}$$

Return to Risk Ratio Agent

Return to Risk Ratio agent (RRR agent) measures the riskiness the stock in previous n trading days by computing standard deviation of the returns to returns ratio and comparing them to the risk appetite of the investor(pre-set). This agent returns signal Bad, Ok, Good.

$$\text{RRR agent } (P_s, t) \rightarrow \{\text{Bad, Ok, Good}\}$$

The data is then fed into various agents and a signal is generated for every trading day (each record).

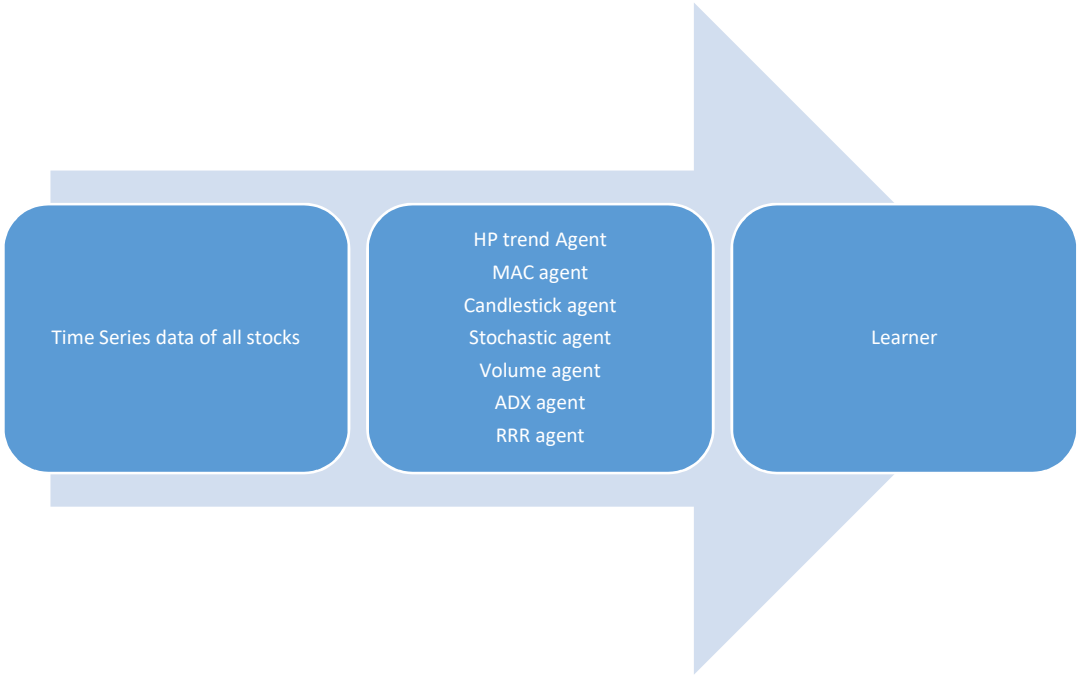
Agent	Signal
Hodrick Prescott Trend Agent	Uptrend, Downtrend, Notrend
Moving Average Crossover Agent	Buy, Sell, Hold
Candlestick Agent	Buy, Sell, Hold
Stochastic Agent	Buy, Sell, Hold
Volume Agent	Strong Volume, Weak Volume
Average Directional Index Agent	Strong Trend, Weak Trend
Return to Risk Ratio Agent	Bad, Ok, Good

These signals are considered as input to the learner.

$$\text{Learner (HP\_trend\_agent, MAC\_agent, Candlestick\_agent, Stochastic\_agent, Volume\_agent, ADX\_agent, RRR\_agent)} \rightarrow \{\text{Buy, Sell, Hold}\}$$

Solution Statement

The solution to build a learner model that will take the output signal of each agent as input and then decide when to Buy, Sell, or Hold on a trade.



The correct decision or actual output of the series will be determined by looking ahead into the series. For next seven days, return generated and standard deviation of daily returns will be calculated. If in next seven days the return to risk ratio is very low (less than pre-defined value) it

would have been better to not get in the trade and correct decision to HOLD. Otherwise if there came a positive return correct decision should have been to BUY and in case of negative return to SELL.

Benchmark Model

As objective of the project is to discover whether the learner can make better decision than all individual indicators the performance of the learner will be compared to performance individual agents namely HP agent, MAC agent, Candlestick agent, and Stochastic Agent.

Evaluation Metrics

F1 score of learner, HP agent, MAC agent, Candlestick agent, and Stochastic Agent will be calculated to compare the results. A substantially better score of learner would indicate the scope of taking this project forward and adding more and more agents to further enhance the capability of the model.

Project Design

Steps involved in the project:

1. The trading data for given time period and given sector will be fetched from yahoo finance and stored on local computer for faster access in future.
2. Initial data check. Only the shares which have trading data for complete time period will be taken so as to remove null values.
3. Data will be fed into agent functions and a signal will be stored for each record.
4. The correct decision will also be worked upon by looking forward into the data.
5. A new file containing just agent signals and correct signals will be generated.
6. This file will be split into 80-20 ratio. The split will be chronological i.e. training data will be from earlier period and testing data will be from later period, rather than random to avoid forward looking bias.
7. Learner algorithm will be applied on training data.
8. F1 score of learner will be compared to same of individual agents to evaluate the model.

Works Cited

investopedia.com. (2017). *Hodrick-Prescott (HP) Filter*. Retrieved from investpedia.com: http://www.investopedia.com/terms/h/hpfilter.asp

investopedia.com. (2017, Jan 1). *Risk*. Retrieved from investopedia.com: http://www.investopedia.com/terms/r/risk.asp

Larsen, J. I. (2010). *Predicting Stock Prices Using Technical Analysis and Machine Learning.* Norwegian Uniersity of Science and Technology.

W. Sean Cleary, H. J. (2014). Market Efficiency. In C. Institute, *2015 CFA Level I Volume 5 Equity and Fixed Income.* Wiley Global FInance.