ML Based Approaches in Stock Monitoring

Name - Amitesh Kumar
Roll-No. 1901023

Problem Description

- The aim of this project is to understand how a weighted stock selection model can affect portfolio performance parameters such as returns, risk and win rate that a portfolio manager could expect to achieve when developing a weighted stock selection model.

- **Stock exchanges are considered major players in financial sectors of many countries. Most.Stockbrokers, who execute stock trade, use technical, fundamental or time series analysis in trying to predict stock prices, so as to advise clients. However, these strategies do not usually guarantee good returns because they guide on trends and not the most likely price. It is therefore necessary to explore improved methods of prediction.**

- A predictive tool that stock brokers can use to guide on exact price movements,as a basis of investment to provide the best advice to their clients,is therefore desirable.So we are using regression to predict the stock.

Attribute Information

The feature/inputs are the weights of the stock-picking concepts as follows

- X1=the weight of the Large B/P concept, X2=the weight of the Large ROE concept, X3=the weight of the Large S/P concept, X4=the weight of the Large Return Rate in the last quarter concept, X5=the weight of the Large Market Value concept, X6=the weight of the Small systematic Risk concept

The outputs are the investment performance indicators (normalized) as follows:

- Y1=Annual Return, Y2=Excess Return, Y3=Systematic Risk, Y4=Total Risk, Y5=Abs. Win Rate, Y6=Rel. Win Rate
- The outputs of UCI Stock Portfolio Performance Dataset is Continuous , so it falls under Regression
- Features : 6 Patterns : 63 in each Period. Total of 4 periods are there

Literature Survey

1. **Using mixture design and neural networks to build stock selection decision support systems.**

   Publisher - Yi-Cheng Liu, I-Cheng Yeh
   Year of Publication - 2015
   Performance measure - In terms of Back test Period using Neural Networks

   Important Points from Survey :

- Procedure of building the stock selection decision support decision :
  Generate weighting combinations of stock-picking concepts with a mixture design
  - Simulate weighting combinations of stock-picking concepts through backtesting
  - Build and analyze the performance prediction model with neural networks
  - Seek the optimal weighting combinations of stock-picking concepts through optimization
  - Validate the optimal weighting combinations of stock-picking concepts through backtesting.
- Dividing dataset according to time frames
- Normalization of performance indicators , all normalized into the same scale(0.2-0.8) during the same time frame 5 years (20 quarters)
- For Four time frames, four neural network models made. K fold validation used to train and test dataset
- Performance of portfolio are Predicted based on return and risk

Literature Survey

Conclusion using model 1 : maximizing return with risk constraints and model 2 : maximizing objective without constraint :

- According to the evaluation of performance prediction models based on neural networks, excess return rates can be predicted more precisely than annualized return rates, total risk than systematic risk, and relative winning rate than absolute winning rate. These may be attributed to the reason that an individual firm's fundamental analysis only effective for stock picking but not useful for market timing
- According to the analysis of performance prediction models based on neural networks, profitable firms' stocks wi relatively cheap prices have higher return rates. The stocks with relatively cheap prices have higher absolute winning rates, lower systematic risks, and lower total risks. Profitable firms' stocks have higher relative winning rates.

2.STOCK PRICE PREDICTION USING ARTIFICIAL NEURAL NETWORKS

- **Publisher -Padmaja Dhenuvakonda , R. Anandan , N. Kumar**
- **Year of Publication - 2020**
- **Method used - Artificial neural networks, RNN and Deep Learning**

Literature Survey

Important Points from the survey :

- Artificial neural networks can be used for stock price prediction and for teaching the network(training)
- Recurrent Neural Networks : Recurrent Neural Networks takes inputs from two types of sources, one is from the present and the other is from the past. Information derived from such sources is to be used for modelling networks' reactions to the newer sets of input data. This is possible with feedback loops wherein output at any instance can be taken as an input to its subsequent instance. This means RNN needs memory. Every input with its huge loads of data needs to be stored in some layers of RNNs.
- Deep Learning is also used
- Dataset is taken from the stock data of a particular company named Infratel. The data set contains information like previous closing, opening, high, low, volume of the stocks of that company. From these datasets, we extract only two months of data; this data will be used to train the model. Using this trained set of data, predicting the 61th day stock market price of that company can be accomplished. The intra-day closing price of stock is given preference as investors have to make a decision on buying with only the stock closing value.
- Conclusion : These models are reasonably efficient in recognizing the patterns that exist in the domain of stock market. This shows that there lies an underlying dynamics, which is very common to all the stock markets.

Literature Survey

3.. Stock Closing Price Prediction using Machine Learning Techniques

- **Publisher : Mehar Vijha , Deeksha Chandolab, Vinay Anand Tikkiwalb, Arun Kumarc**
- **Year of Publication : 2019**
- **Method used : Artificial Neural Networks**

Important Points :

- The historical data for the five companies has been collected from Yahoo Finance . The dataset includes 10 year data from 4/5/2009 to 4/5/2019 of Nike, Goldman Sachs, Johnson and Johnson, Pfizer and JP Morgan Chase and Co. The data contains information about the stock such as High, Low, Open, Close, Adjacent close and Volume. Only the day-wise closing price of the stock has been extracted.
- Six new variables have been created for the prediction of stock closing price. These variables have been used to train the model.
- 1. Stock High minus Low price (H-L) 2. Stock Close minus Open price (O-C) 3. Stock price’s seven days’ moving average (7 DAYS MA) 4. Stock price’s fourteen days’ moving average (14 DAYS MA) 5. Stock price’s twenty one days’ moving average (21 DAYS MA) 6. Stock price’s standard deviation for the past seven days (7 DAYS STD DEV)

Literature Survey

Important Points Cont.:

- Ann used - The model works with three layers. It consists of input layer, hidden layer and the output layer. The input layer consists of new variables which are H-L, O-C, and 7 DAYS MA, 14 DAYS MA, 21 DAYS MA, 7 DAYS STD DEV and Volume . The weights on each input load is multiplied and added and sent to the neurons. The hidden layer or the activation layer consists of these neurons. The total weight is calculated and is moved to the third layer which is the output layer. The output layer consists of only one neuron which will give the predicted value in terms of closing price of the stock.
- Results : The comparative analysis based on RMSE, MAPE and MBE values clearly indicate that ANN gives better prediction of stock prices as compared to RF. Results show that the best values obtained by ANN model gives RMSE (0.42), MAPE (0.77) and MBE (0.013).
- Conclusion : The historical dataset available on company's website consists of only few features like high, low, open, close, adjacent close value of stock prices, volume of shares traded etc., which are not sufficient enough. To obtain higher accuracy in the predicted price value new variables have been created using the existing variables. ANN is used for predicting the next day closing price of the stock

4. Study of Machine learning Algorithms for Stock Market Prediction

- **Publisher : Ashwini Pathak,Sakshi Pathak**
- **Year of Publication - 2020** **- Method used - Logistic regression,KNN Clustering,SVM**

Results

Period 2 -Using Stochastic Gradient Descent for Linear Models With hyperparameter Tuning on Dataset

- **Annual Return** - MSE 0.008 , **Hyper Parameters** : **Learning rate**- 0.1,**Rho**-0..0001,**Max Itr**-2000
  - R2 score - 0.785
- **Excess Return** - MSE 0.005, **Hyper Parameters** : **Learning rate**- 0.1,**Rho**-0..0001,**Max Itr**-2000
  - R2 Score - 0.420
- **Systematic Risk** - MSE 0.007 , **Hyper Parameters** : **Learning rate**- 0.1,**Rho**-0..0001,**Max Itr**-1000
  - R2 Score - 0.470
- **Total Risk** - MSE 0.003 , **Hyper Parameters** : **Learning rate**- 0.001,**Rho**-0..0001,**Max Itr**-1000
  - R2 Score - (-0.608)
- **Abs. Win Rate** - MSE 0.007 , **Hyper Parameters** : **Learning rate**- 0.001,**Rho**-0..0001,**Max Itr**-1000
  - R2 Score - 0.189
- **Rel. Win Rate** - MSE 0.013 , **Hyper Parameters** : **Learning rate**- 0.1,**Rho**-0..0001,**Max Itr**-1000
  - R2 Score - 0.473
- We had decent MSE scores, that means our model fits pretty good.

Results

Period 1 -Using K fold Cross Validation on Dataset and checking MSE of Training and Test Set
Note : K=5,Values are approximated upto 3 decimal places

- **Annual Return** - MSE Train set - 0.006+-0.0006 MSE Test Set - 0.009+-0.0033
- **Excess Return** - MSE Train set - 0.007+-0.0006 MSE Test Set - 0.011+-0.0043
- **Systematic Risk** - MSE Train set - 0.007+-0.0008 MSE Test Set - 0.012+-0.0057
- **Total Risk** - MSE Train set - 0.006+-0.0010 MSE Test Set - 0.012+-0.00703
- **Abs. Win Rate** - MSE Train set - 0.0104+-0.0011 MSE Test Set - 0.015+-0.0062
- **Rel. Win Rate** - MSE Train set - 0.008+-0.0009 MSE Test Set - 0.011+-0.0044
- We had decent MSE scores, that means our model fits pretty good.

Results

Period 3 -Using K fold Cross Validation on Dataset and checking MSE of Training and Test Set
Note : K=5,Values are approximated upto 3 decimal places

- **Annual Return** - MSE Train set - 0.003+-0.0005   MSE Test Set - 0.005+-0.0031
- **Excess Return** - MSE Train set - 0.002+-0.0003   MSE Test Set - 0.004+-0.0020
- **Systematic Risk** - MSE Train set - 0.007+-0.0016   MSE Test Set - 0.012+-0.0095
- **Total Risk** - MSE Train set - 0.006+-0.0013   MSE Test Set - 0.009+-0.0074
- **Abs. Win Rate** - MSE Train set - 0.0103+-0.0014   MSE Test Set - 0.018+-0.0091
- **Rel. Win Rate** - MSE Train set - 0.007+-0.0011   MSE Test Set - 0.010+-0.0054

Results

Period 4 -Using K fold Cross Validation on Dataset and checking MSE of Training and Test Set
Note : K=5,Values are approximated upto 3 decimal places

- **Annual Return** - MSE Train set - 0.003+-0.0004 MSE Test Set - 0.004+-0.0025
- **Excess Return** - MSE Train set - 0.002+-0.0002 MSE Test Set - 0.034+-0.0016
- **Systematic Risk** - MSE Train set - 0.001+-0.0002 MSE Test Set - 0.002+-0.0011
- **Total Risk** - MSE Train set - 0.001+-0.0001 MSE Test Set - 0.002+-0.00118
- **Abs. Win Rate** - MSE Train set - 0.008+-0.0013 MSE Test Set - 0.015+-0.0086
- **Rel. Win Rate** - MSE Train set - 0.008+-0.0004 MSE Test Set - 0.012+-0.0021

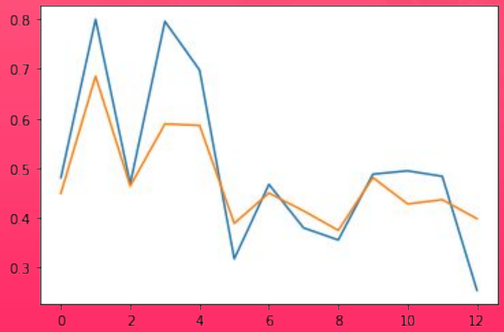
Analysis of Results

- Since the outputs are continuous we can use Regression Algorithms to train and test the model. So we have used Linear Regression for this purpose.
- Since there are 6 features (X) or weights of stocks, After observing each output with all the features , We came to conclusion that we need to use Multivariate Linear regression because there was no linear relation seen between individual features and Independent outputs.
- In Multivariate Linear Regression there are more than one input variables used to estimate the target.
- Used k Fold cross validation and Stochastic gradient descent approach for analysis of dataset
- In SGD , Model was trained epoch by epoch until we got very less avg loss and using the hyperparameter tuning chose the best ones. The results were reported
- In K Fold , We trained the model for k-1 folds and tested it for other. The train and test MSEs are reported in the slides.

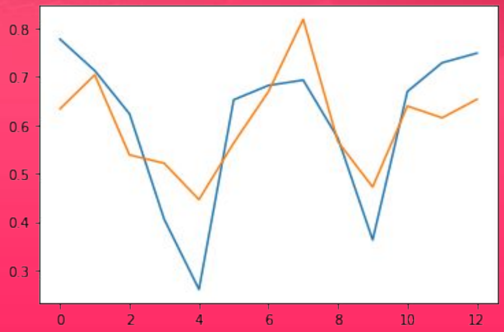
Analysis of Results

Actual vs Predicted Values using SGD

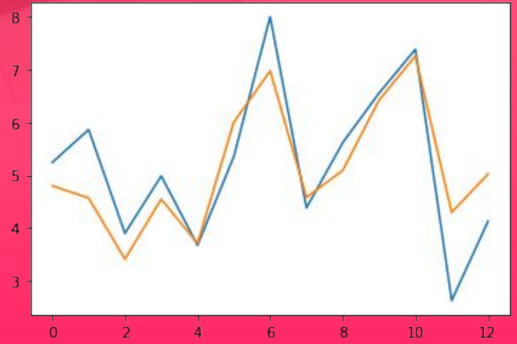
Annual Return



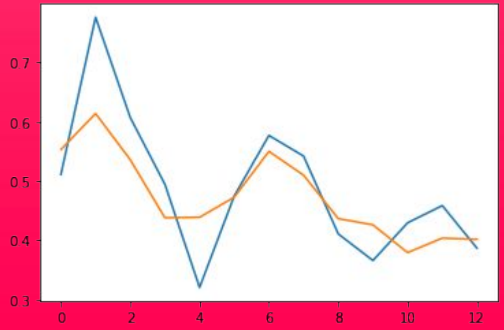
Excess Return



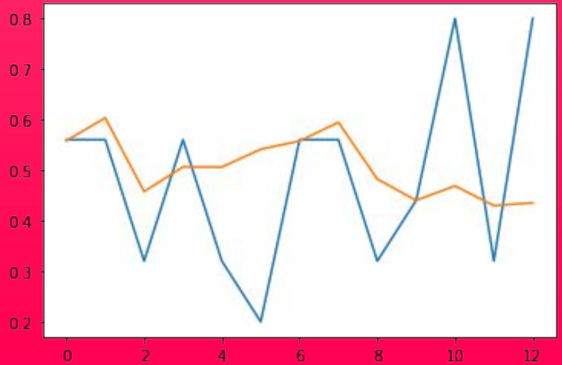
Systematic Risk



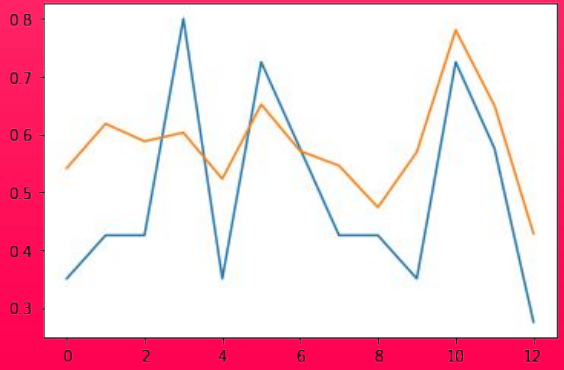
Total Risk



Abs. Win Rate



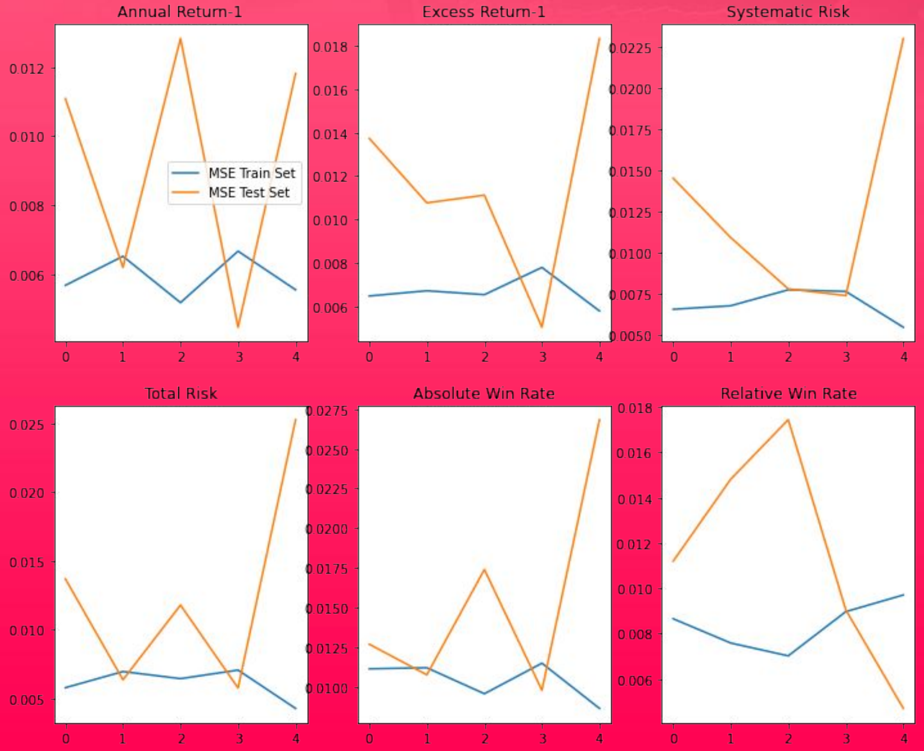
Rel. Win Rate



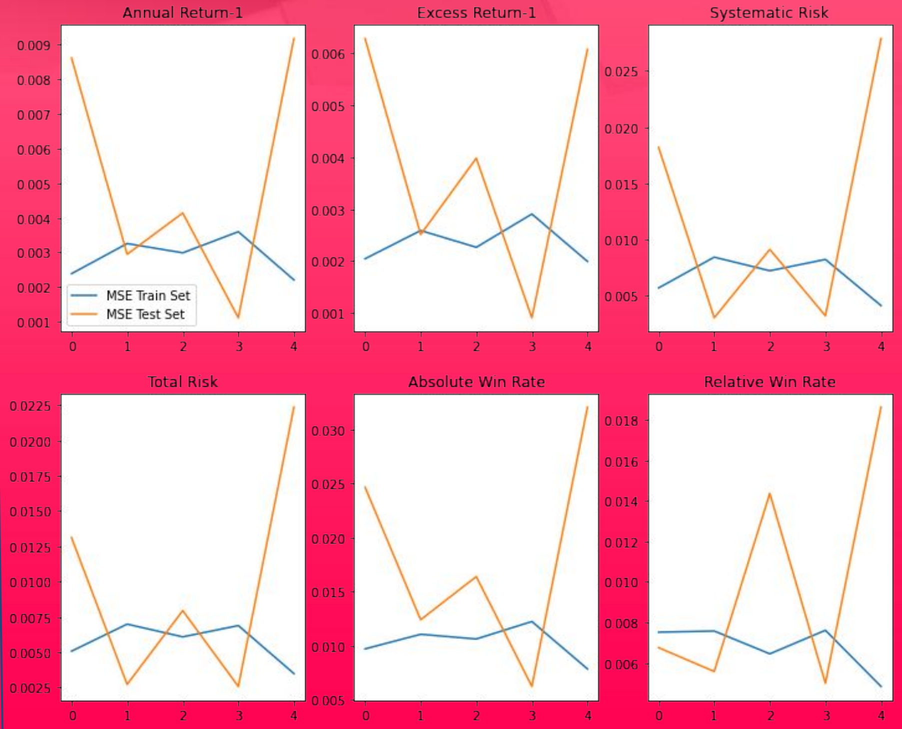
Analysis of Results

Graphs for MSE OF TRAINING SET VS TEST SET for all 5 folds USING K FOLD

Period 1



Period 3



Conclusion

- In case of SGD, In the above graphs and results we can see that the errors are very less, showing the actual and the predicted values very similar in nature. So using multivariate linear regression works fine.

- In case of K-Folds, again we in the above graphs and Results , it is clear that the MSE of test set is very low and comparable to training set so the model is performing well with this method. Thus using multivariate linear regression works.

- R2 score is negative only when the chosen model does not follow the trend of the data, so fits worse than a horizontal line. In our results we are getting some r2 scores where one of them does not fits the model.Rest are performing somewhat good.

- R2 score and MSE results are best for Annual Returns which means the model fits well with it and R2 score and MSE results for Absolute win rate are worst .

Thank you !