Prakhar Tripathi

(MLND Capstone Project)

Feb ,2019

Domain Background-

Being an economics lover I always had a fascination towards stock markets round the world. Also I decided to pursue my MLND project on a topic regarding this. I have studied various stock predictors in stock market which us ML to predict all about a particular company or stocks of the day. There is always rise and fall in stock prices along the day. I have a personal interest in AMD (Advanced Micro Devices, Inc) which is a microprocessor company and also a GPU provider of my system.Also due to unpredictable stock market the use of ML is much necessary to stop from tragedy of 2008 markets crash.

Problem statement

The project deals with prediction of a particular stock of NASDAQ where its performance will firstly be understood and later decisions will be made on near future of stocks. This would provide customer idea to invest or not.

This project is designed using neural networks so that there a continuous Update in stock ups and down. The performance is analyzed using long study of stock market.

Even with new libraries like tensor flow developed by Google can easily solve these stuff by easily working on heavy data sets. I would like to model stock prices correctly, so as a stock buyer you can reasonably decide when to buy stocks and when to sell them to make a profit. This is where time series modelling comes in.

I shall be using time-series model known as Long Short-Term Memory. LSTM models are powerful, especially for retaining a long-term memory. to predict the price of my favourite stock AMD.

Data sets and inputs-

I have used data of set NASDAQ from Kaggle(https://www.kaggle.com/gunhee/amdgoogle/version/1)to this project.The project is able to predict everyday final price of AMD stocks after rigorous training and finally testing.
Data pre- processing----
1.Data discretization
2.Data integration
The data sets used completely continuous because numeric variables that have an infinite number of values between any two values.Also a continuous variable can be numeric or date/time. For example, the length of a part or the date and time a volume or stock point is received.
There are 2336 values provided by kaggle in AMD data set from may 2009 to september 2018.   The tables include volumes of stocks each day,opening and closing points per share.

Solution statement-

The main solution will be found using LSTM because LSTM`s are very powerful in sequence prediction problems because they're able to store past information. This is important in our case because the previous price of a stock is crucial in predicting its future price.
The major output will be in Jupyter notebook (.ipynb format)and will be supported using various python modules.
Libraries used for LSTM from keras

```
yrecurrent_activation='hard_sigmoid', use_bias=True,
kernel_initializer='glorot_uniform', recurrent_initializer='orthogonal',
bias_initializer='zeros', unit_forget_bias=True, kernel_regularizer=None,
recurrent_regularizer=None, bias_regularizer=None, activity_regularizer=None,
kernel_constraint=None, recurrent_constraint=None, bias_constraint=None, dropout=0.0,
recurrent_dropout=0.0, implementation=1, return_sequences=False, return_state=False,
go_backwards=False, stateful=False, unroll=False))
```

Libraries like pandas ,matplotlib, numpy also used.
The final prediction shall be on comparing actual and estimated stock points.
The basic output I would love love to get prediction of share point of day later on using the results of training data sets.Also it would be great to see the volumes growth of AMD company.

Benchmark models-

The model will be actually studied using continuous data so it is studied for final prediction using Linear regression.Optimization and performance will be studied using recurrent neural network models.
The LSTM have back propagation with memory storage prowess allowing it to handle large data sets.The neural net model is primary model for me where model be re evaluated again to make final prediction.

Metrics used-

For analyzing model various metrics used are-
1. RMSE(root mean squared evaluation)-It is very good model for large continuous data.Unlike the absolute mean error it severely punishes large Error.
It is-

$$((Y_{pred}-Y_{obs})^2/n))^{1/2}=Error$$

Y_{pred}-linear regression output
Y_{obs}-Neural network output
n - total input
d/d(E)=gradient descent

Project Design-

1. Data sets-a.Data set of AMD from kaggle.
    b .use pandas to import data set.
        eg-d=pd.read_csv('Filename)
    c . Analyzing data sets from top and tail.
    d . split dat in 3:7of testing: train
2. Develop primary and secondary models- LSTM(primary)and linear regression secondary model.Curves will be plotted to increase the visualization using graphs.
3. Libraries used -Keras,numpy,pandas ,matplotlib,scikit learn etc.

References-

1. https://www.kaggle.com/gunhee/amdgoogle/version/1
2. https://ieeexplore.ieee.org/abstract/document/23959/metrics#metrics
3. https://www.kaggle.com/johanvandenheuvel/lstm-model-of-stockdata/comments
4. https://www.quantcon.com/
5. https://www.kdnuggets.com/2018/11/keras-long-short-term-memory-lstm-model-predict-stock-prices.html

6. https://www.datacamp.com/community/tutorials/lstm-python-stock-market