# Financial Prediction, Some Pointers, Pitfalls, and Common Errors

Presented by: Jeevesh Jain

# Index

# Introduction

- Financial Prediction is focused on due to high profit chance.
- There is a justiable scepticism surrounding the idea that it is possible to make money by predicting price changes in a given market based only on its past behaviour and a number of publicly available indicators
- The efficient market hypothesis (EMH), alternatively known as the efficient market theory, is a hypothesis that states that share prices reflect all information and consistent alpha generation is impossible.

# Demerits(Merits?) of a perfect prediction

- The efficient markets hypothesis will no longer apply to the person with the model until everybody gains access to the same technology and things even out once more. Given that the efficient markets hypothesis relies on perfect knowledge; perfect prediction technology would only serve to enforce the conditions under which that technology is useless; as soon as a possible profit is predicted, it is snapped up and expected profits return to the level of the risk free return plus a risk premium associated with a stock holding.
- Chances of superior profits occur when brokers set their prices incorrectly and investors are able to spot the discrepancy before it is corrected.
- Perfect prediction for all would remove these discrepancies and with them the opportunities for superior prot.

# Time Series Prediction

- **Time series forecasting** is the use of a model to **predict** future values based on previously observed values. **Time series** are widely used for non-stationary data, like economic, weather, stock price, etc.
- The two main neural network based approaches to time series forecasting discussed in this paper are time windowing and recurrent networks.
- What is lag?

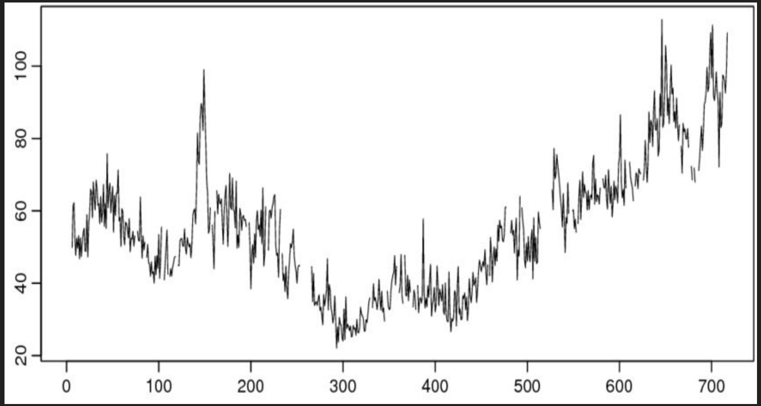
  Lag in common terms is delay, it specifies how much our time series is delayed by so that we get many/one step ahead prediction.

- What is Autocorrelation?

  Autocorrelation is the measure of similarity of the time series with itself. Autocorrelation is a major factor which is used to determine the minimum lag required for computations.

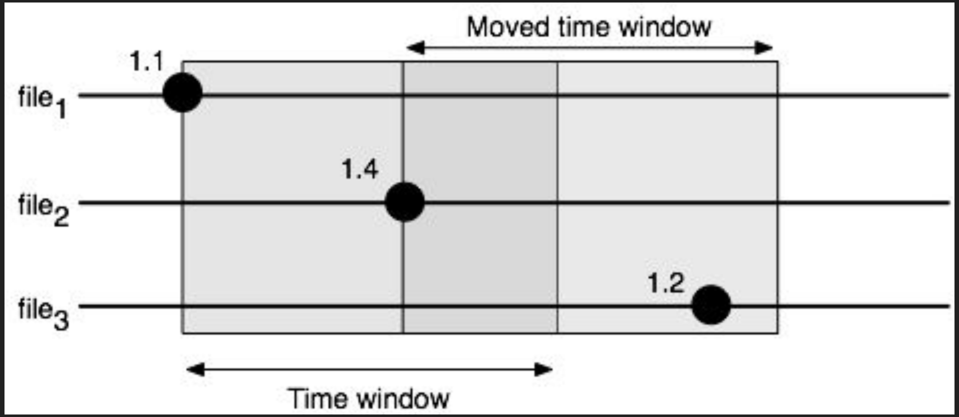
# Univariate vs Multivariate time series

- **Univariate Time Series**: These are datasets where only a single variable is observed at each time, such as temperature each hour.
- **Multivariate Time Series**: These are datasets where two or more variables are observed at each time.



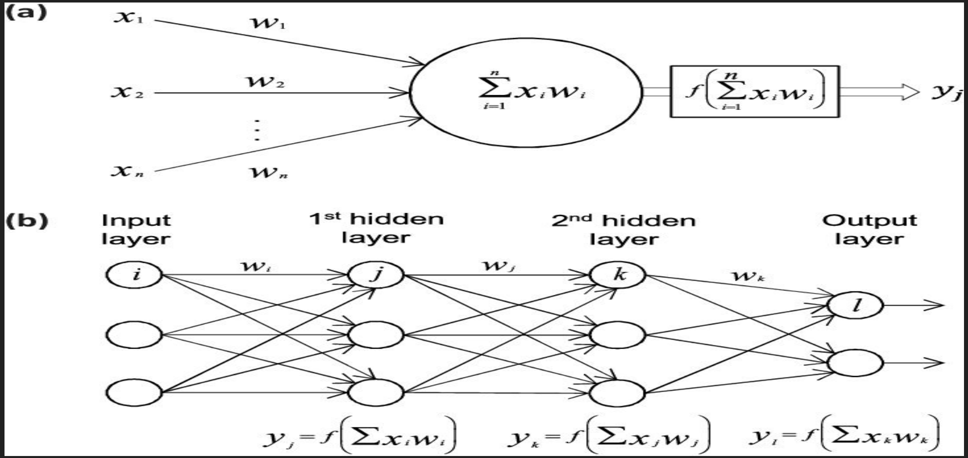

# Time Windowing

- Converting temporal dimensions to spatial vectors by segmenting them according to a particular pattern.
- There are two types of time windowing, fixed time windowing and rolling/sliding time windowing.
- 



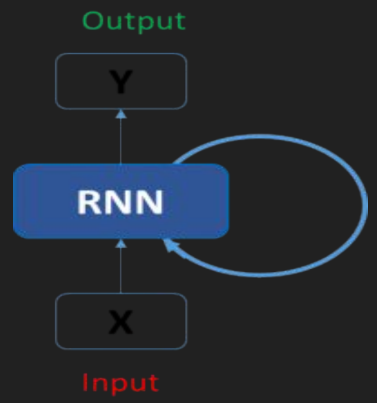
# Recurrent Networks or RNNs(1)

- A neural network are sore of composite functions which are universal function approximators.
- Basic 3 layers of NN: input,hidden and output layer.
- Each of the layers are comprised of nodes which in turn are connected to other layers through weights.
- These weights are first arbitrarily selected and then re-selcted via the predicted output vs expected output comparison through backpropagation.



# Recurrent Networks or RNNs(2)

- [Elman, 1990] showed how an otherwise feedforward network with a recurrent context layer which took a copy of the network's hidden layer at time t-1 and re-applied it in addition to the input vector at time it was able to learn temporal dependencies.
- At all the time steps weights of the recurrent neuron would be the same since its a single neuron now. So a recurrent neuron stores the state of a previous input and combines with the current input thereby preserving some relationship of the current input with the previous input. This is how LTSM is able to store some memory within themselves for some time!



# How do RNNs/LSTMs(Long Short Term Memory Networks)  memorize?

- LSTMs build short and long term memories by revealing data to the hidden nodes in a sequential fashion.
- Each LSTM node is comprised of LSTM cells.
- The long term and short term memory is updated in each LSTM cell upon being exposed to each subsequent element of the sequence , conditional on the output of each ‘gate’.

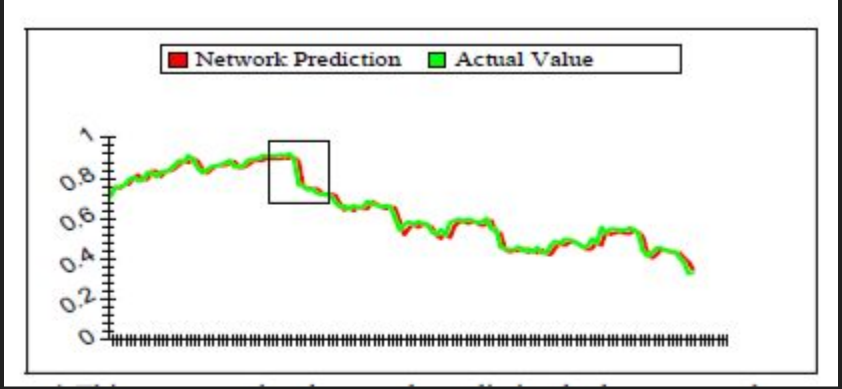
# One Step Ahead Prediction vs Many Step Ahead Prediction

One step Ahead prediction predicts for the next value of time, (in accordance to the time difference series is being subjected to) whereas multi-step ahead prediction can predict many more steps depending upon the lag provided to the data and the depth and recurrence function of the RNN.

- [Refenes, 1991] used overlapping time windows to convert exchange rate series into spatial vectors. The data [hourly updates of 260 days worth of US dollar/DM exchange rate values] were coded into two moving windows with the intention of mapping the contents of the earlier window of size n onto those of the second, later, window of size m. The two windows move across the data series at a step of s. The task then is to predict the contents of window m from those of n. The authors show how important network design is; choosing the sizes of n;m and s carefully as well as constructively adding hidden units to maximise the data t. One step ahead prediction was achieved by setting m = 1, i.e setting the output vector to be a single value. Multi step prediction was achieved by taking the single valued prediction and feeding it back as input rather than by extending the size of the output window, m.
-

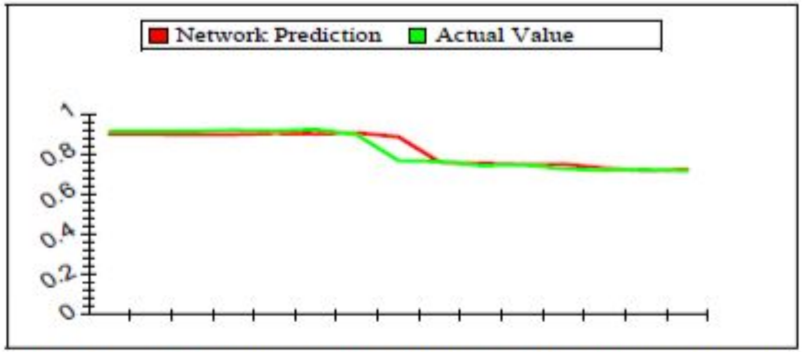
# Common Pitfalls to Avoid (One step Ahead)

- The current values are shown as the output ( at t=t+1) is being predicted the same as what it is right now and a “too good to be true” output is obtained.



# Common Pitfalls to Avoid (One Step Ahead)

- Future is not being predicted, instead current values are being predicted but at a different point of time, making us lose the chances of profits.




b) A closer look at the area in the box in a) shows how the chance of a profit is missed.

# Common Pitfalls to Avoid (Many Step Ahead)

- When predicting several steps ahead, start the prediction from many di erent steps along the time series and check to see that there is no coincidence in your results. An example is a predictor which simply follows the current or overall trend and so appears to be correct for several steps.Only when the predictor consistently forecasts turning points can you claim success.
- This can be done by increasing the number of iterations per step i.e epochs which reduces overfitting and chances of coincidence are minimalized.
- We also sometimes do a mistake of not accounting the full process's profit and determine whether a transaction is successful or not by seeing a single step.

# Why can’t we use some linear ML model instead of RNNs?

- OLS or Ordinary Least Squares (Linear Regression) cannot be used for the temporal dataset

| Properties of Time Series Data | Meaning? | Can OLS deal with this? | Why/Why Not? |
|---|---|---|---|
| Sequentially ordered | Outcome today depends on outcome yesterday | No | • Order of the data does not matter<br>• OLS requires no correlation between observations of the error term (no autocorrelation) |
| Displays trends | Different data series may show high correlation even if their non-trend component is not correlated (i.e. spurious correlation) | No | • OLS requires no perfect correlation between independent variables (no multicolinearity) |
| May not be stationary | Mean of the data changes over time | No | • OLS requires that the variance of error does not change for each observation/ a range of observations (no heteroskedasticity) |

Source: Author compiled ; *Basic Econometrics* ,Damodar N. Gujarati ; CMU Lecture notes

# Another method (suggested): CNN

- A Convolutional Neural Network can extract information from the temporal structure of the data by preserving the temporal structure of the time series in the input layer and using filters which can look for patterns to extract features. This gives us two advantages; we indirectly use a multivariate approach to train the model and CNN being a lighter model than RNN, works faster. The results are not compromised to a huge extent and in many cases, CNNs turn out to be performing better than LTSMs.

# Why to choose CNNs and LTSMs for time series prediction afterall?

However, both CNNs and LSTMs can “learn an internal representation of the time series data and ideally achieve comparable performance to models fit on a version of the dataset with engineered features” How?

- The input layer preserves the temporal structure of data
- The hidden layer/s use complex computational nodes: The different computational nodes used in a CNN vis-à-vis a LSTM network result in the two type of networks processing the data differently.

# Avoiding False Assumptions

- A common rookie mistake done by people is to blindly believe the model they have formulated. Such mistakes can be avoided by a few ways.
- One involves reducing the number of hidden units, or even using a perceptron on the same training task. If you have looked at your data set and decided that N hidden units would produce a good t, and found that it seems to do so, then a lot can be learned by the deterioration (or lack of it) a network displays when N is reduced.
- Another way is to test the network on data with a totally unrelated structure; a sine wave for example. If you still seem to be getting good results, then it is clear that there is an error in your interpretation

# Market Modelling: Taking in account the trade costs and transaction frequencies

- You may be able to predict one step ahead every minute and accumulate sixty small profits an hour, but if you have to pay 0.05% on each transaction, your sum may actually dwindle to nothing. Similarly it is not possible to trade at such a speed and nor is it so easy to swap from a buying to a selling position at will without accounting for differences in brokers' buy and sell prices.

# Bibliography

## The slides have excerpts from:


- On Developing a Financial Prediction System: Pitfalls and Possibilities(http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.2708&rep=rep1&type=pdf)
- The Efficient Market Hypothesis (https://www.princeton.edu/~ceps/workingpapers/91malkiel.pdf)
- Financial Performance evaluation of Construction Industries (http://www.ijsrp.org/research-paper-0117/ijsrp-p6129.pdf)
- One-step and multi-step ahead stock prediction using backpropagation neural networks(https://www.researchgate.net/publication/271556009_One-step_and_multi-step_ahead_stock_prediction_using_backpropagation_neural_networks)
- Stock Market Prediction using LTSM Recurrent Neutral Network(https://reader.elsevier.com/reader/sd/pii/S1877050920304865?token=18384927099BB2281AB654A2C02B025526B4640ADB6C46D9E63E579505E9E6141FC5C9A49D9433C4EC5A22B5F17B8C24)

  The elements of Statistical Learning(https://web.stanford.edu/~hastie/Papers/ESLII.pdf)