PROJECT OF Kevin Rodrigues

Assignment 4

Paste Assignments 1 - 3 here as is, except in Assignment 3, respond to each of your facilitator’s comments within each of their comments. As always, retain these gray parts. Compared with the previous assignment, text added below is underlined. Deletions are shown.).

# Assignment 3: Project Design, version 1

Keep in mind the evaluation matrix at the end as you do the work and use it to guide what you submit. Use no more than 6 pages of 12-point text excluding figures. You may include as many appendices as you wish for reference. Parts of these may be read as needed.

## 3.1 Final Requirements

List your final requirements, numbering them in the form DiX and NiX where:

D/N means “Definite” / “Nice to do” (two categories, not three)

i = 1, 2, 3, …

X=L and the goal is a *learning* goal – or – X=F and the goal is *functional*

You will reference these numbered requirements in the rest of the term, when you will be asked to show what the project accomplished.

D1L = Dataset of a particular stock

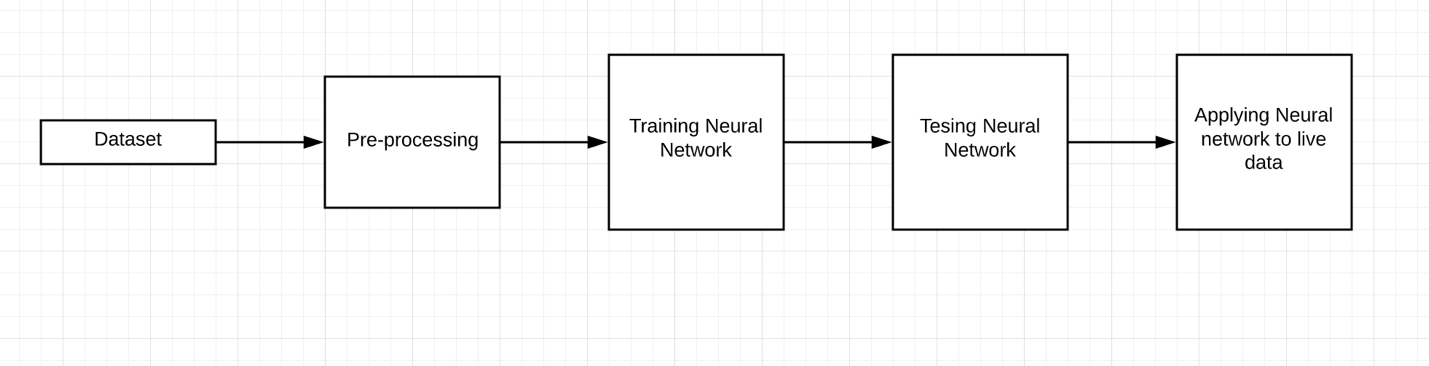
D2F = Comparison between different neural network models such as MLP and LSTM on predicting closing prices for a particular stock.[1][2]

N1L = Live Data of a particular stock for checking accuracy on live data.

N2F = Creating a model that can predict if the week is good or bad based on a rule. (for eg. Avg closing price for 1 week is greater than x then it is a good week.)

## 3.2 Design and Theory

Describe the design of your proposed system. Use annotated diagrams. Explain the theory behind your design. Explain how the two technologies will interface or compare. The reader should understand how you plan to fit the pieces together. Show this at a high level, as well as providing as much detail as you can at this point. Include at least one (meaningful) figure, for example, using boxes and arrows showing data flow.



Process for Training a Neural Network for Stock Prediction

## 3.2.1 Dataset:

## We can get the data from websites like yahoo finance:

<https://finance.yahoo.com/quotes/OCR,dataset/view/v1/>

We can also get the data using the yahoo finance API. [3]

## 3.2.1 Data Pre-processing:

In this step we need to Normalize the data using methods such as Z-score, Min-Max[4], Splitting

Data into train and test. Creating new labels such as good week bad week.

## 3.2.2 Training Neural Network:

## Creating a Neural Network to train on the time series data of the stock and provide a prediction.

## 3.2.3 Testing Neural Network:

We can test the Neural Network by predicting the target variable using test set.

We can then optimize the neural network i.e we can try out different optimizers.

## 3.2.4 Applying Neural Network to Live Data:

This can be used for real world purposes and to check if the model is good enough to be implemented in real world.

## 

## 3.3 Tools

Describe the tool(s) you will probably use, or explain why you will build from scratch. It is OK if you say "I will use tool 1 or tool 2." Support the fact that you have reasonably investigated and tried out tools. Explain your choice. Show samples that make you and us reasonably confident of your choices. Show that you understand how the tools work.

TensorFlow and Keras:

I would like to build my model using the TensorFlow and Keras framework. Keras has many models that are pre-trained. I would Like to build my own neural network to obtain results and compare it to a model that I will train using transfer learning in Keras[5][6]

Pytorch:

Pytorch is similar to Keras. It is an implementation by Facebook.[7]

## 3.4 Risk Retirement

Identify and prioritize the 5 top risks in carrying out the project. Try as best you can to retire (set to rest) the top two by the time you submit this, by means of experiments, prototypes, or work-arounds. Explain how you did this. Explain how you will retire the remaining ricks in advance.

**Stock Market Uncertainty:**

Stock Market is uncertain since it is affected by factors such as Politics, Natural Calamities, etc.

*Work-arounds:*

This risk is very difficult to mitigate. You will need multiple models such as predicting weather.

**LSTM risks:**

LSTM works in a way that it remembers some of the previous data. You can provide this as an input to the initialization function. This number is usually decided by looking at trends in the data. But if this trend is not followed for some part of the training data the predictions can be affected.

*Work-around:*

We need to try to get as much data as possible and try and set this parameter wisely.

## 3.5 Schedule

Explain in outline the steps you intend to take to carry out the project. Show the completion of the stages. Include a schedule, as detailed as can be reasonably foreseen.

Week 4: Implementing basic MLP

Week 5: Implementing LSTM and comparing

Week 6: Trying to implement Nice to have goals

## 3.6 References

[1] <https://towardsdatascience.com/the-fall-of-rnn-lstm-2d1594c74ce0>

[2] <https://machinelearningmastery.com/how-to-develop-multilayer-perceptron-models-for-time-series-forecasting/>

[3] <https://blog.rapidapi.com/how-to-use-the-yahoo-finance-api/>

[4] <https://www.geeksforgeeks.org/data-normalization-in-data-mining/>

[5] <https://machinelearningmastery.com/multivariate-time-series-forecasting-lstms-keras/>

[6] <https://www.tensorflow.org/beta/tutorials/text/time_series>

[7] <https://medium.com/@UdacityINDIA/tensorflow-or-pytorch-the-force-is-strong-with-which-one-68226bb7dab4>

# Assignment 4: Project Design Plus, version 2

Keep in mind the evaluation matrix at the end as you do the work and use it to guide what you submit. Use no more than 6 pages of 12-point text excluding figures. You may include as many appendices as you wish for reference. Parts of these may be read as needed. This revision is your final view of the design prior to implementation (though you may still change it when you implement).

## 4.1 V2 Final Requirements

List your final requirements, refined again if necessary, numbering them in the form DiX and NiX where:

D/N means “Definite” / “Nice to do” (two categories, not three)

i = 1, 2, 3, …

X=L and the goal is a *learning* goal – or – X=F and the goal is *functional*

You will reference these numbered requirements in the rest of the term, when you will be asked to show what the project accomplished.

D1L = Dataset of a particular stock

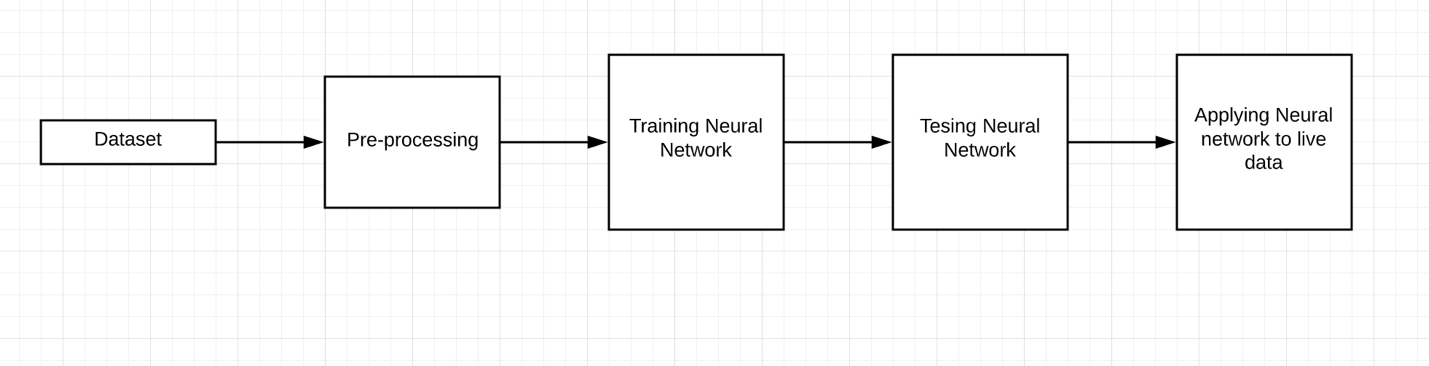
D2F = Comparison between different neural network models such as MLP and LSTM on predicting closing prices for a particular stock.[1][2]

N1L = Live Data of a particular stock for checking accuracy on live data.

N2F = Creating a model that can predict if the week is good or bad based on a rule. (for eg. Avg closing price for 1 week is greater than x then it is a good week.)

## 4.2 V2 Design and Theory

Describe the final pre-implementation version of the design of your proposed system. Use annotated diagrams. Explain the theory behind your design. Explain how the two technologies will interface or compare. The reader should understand how you plan to fit the pieces together. Show this at a high level, as well as providing as much relevant detail as you can ~~at this point~~. Include at least one (meaningful) figure.



Process for Training a Neural Network for Stock Prediction

## 4.2.1 Dataset:

## We can get the data from websites like yahoo finance:

<https://finance.yahoo.com/quotes/OCR,dataset/view/v1/>

We can also get the data using the yahoo finance API. [3]

dataset has the date, open, high, low, closing, Adj closing, volume features.

## 4.2.1 Data Pre-processing:

Our predictors would be the open, high, low and volume. And our target will be Closing price.

We need to Normalize the predictor data using methods such as Z-score, Min-Max[4].Splitting Data can be done using time frames. Suppose we have data for 5 years we can use data of 4 years for training and data of last one year for testing. This will help us to split the data as well as capture trends in the time series data. Creating new labels such as good week bad week.

## 4.2.2 Training Neural Network:

Creating a Neural Network to train on the time series data of the stock and provide a prediction. Our predictors would be the open, high, low and volume. And our target will be Closing price.

We can have the first dense layer with as many perceptron’s as features needed to train the data.

Then we can have several Hidden layers. Finally we shall have a layer with a single perceptron.

We shall use Relu as the activation function throughout all the layers. We can try out different optimizers such as Adam, Sgd, Adagrad, etc.

## 4.2.3 Testing Neural Network:

We can test the Neural Network by predicting the target variable using test set. Since stock is unpredictable and we are doing a regression prediction accuracy is of no use and we can judge the model based on loss functions such as Mean Squared Error, Cross Entropy, etc.

We can then optimize the neural network i.e we can try out different optimizers.

## 4.2.4 Applying Neural Network to Live Data:

This can be used for real world purposes and to check if the model is good enough to be implemented in real world. This model can be used to find trend in stock rather than to predict exact stock value price. It can be used to see if we should invest or sell.

## 4.3 V2 Tools

Describe the tool(s) you will definitely use, or explain why you will build from scratch. ~~It is OK if you say "I will use tool 1 or tool 2."~~ Support the fact that you have reasonably investigated and tried out tools. Explain your choice. Show samples that make you and us reasonably confident of your choices. Show clearly that you understand how the tools work (in this, the next section will help too).

TensorFlow and Keras:

I would like to build my model using the TensorFlow and Keras framework. Keras has many models that are pre-trained. I would Like to build my own neural network to obtain results and compare it to a model that I will train using transfer learning in Keras[5][6]

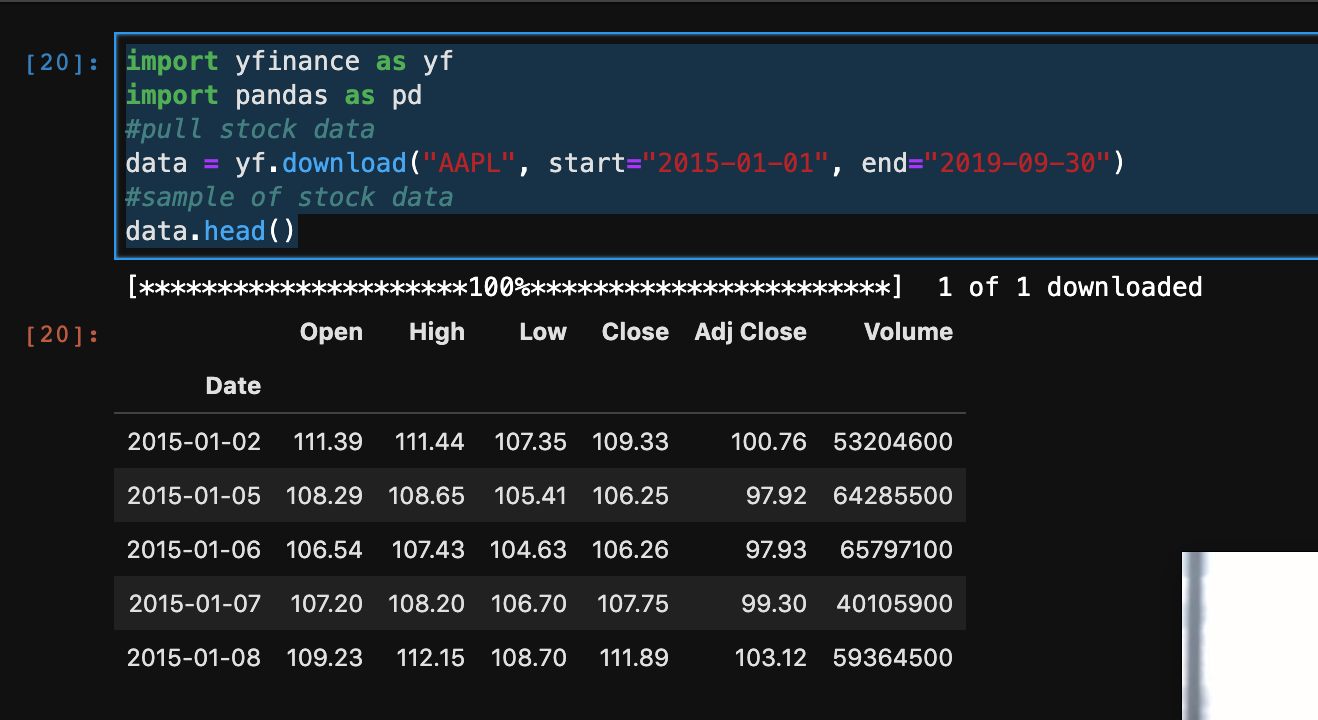
Pytorch:

Pytorch is similar to Keras. It is an implementation by Facebook.[7]

## 4.4 Implementation Fragments

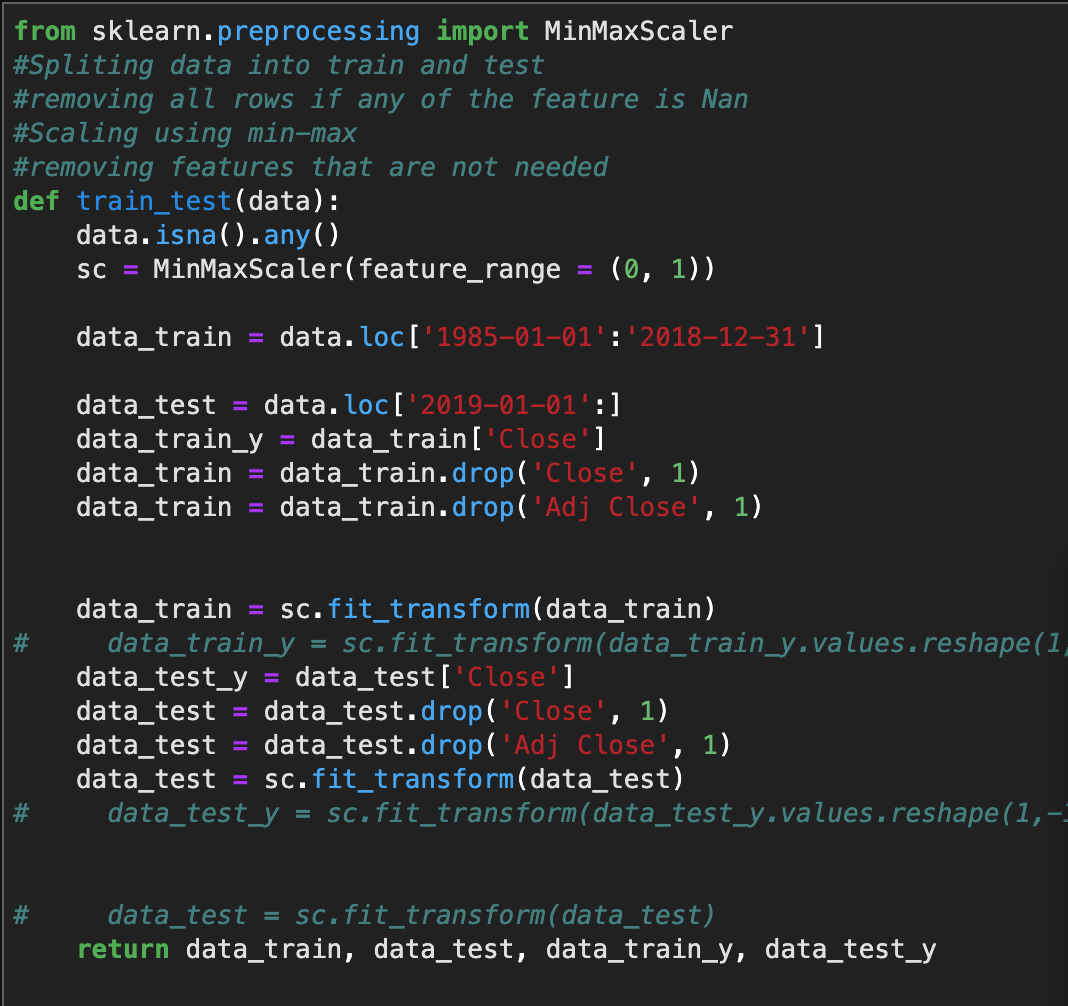
Show enough *parts* of an implementation—or a simplified form of it—to convince yourself and the reader that you will have the implementation of the definite requirements completed on time. Try to make these part of the actual intended application, but they can be experimental or exploratory in nature. Your choices can coordinate with section 4.4 below. Cut and paste commented code below.

Dataset Download:



I used the Yahoo Finance API to get the Apple Stock Data

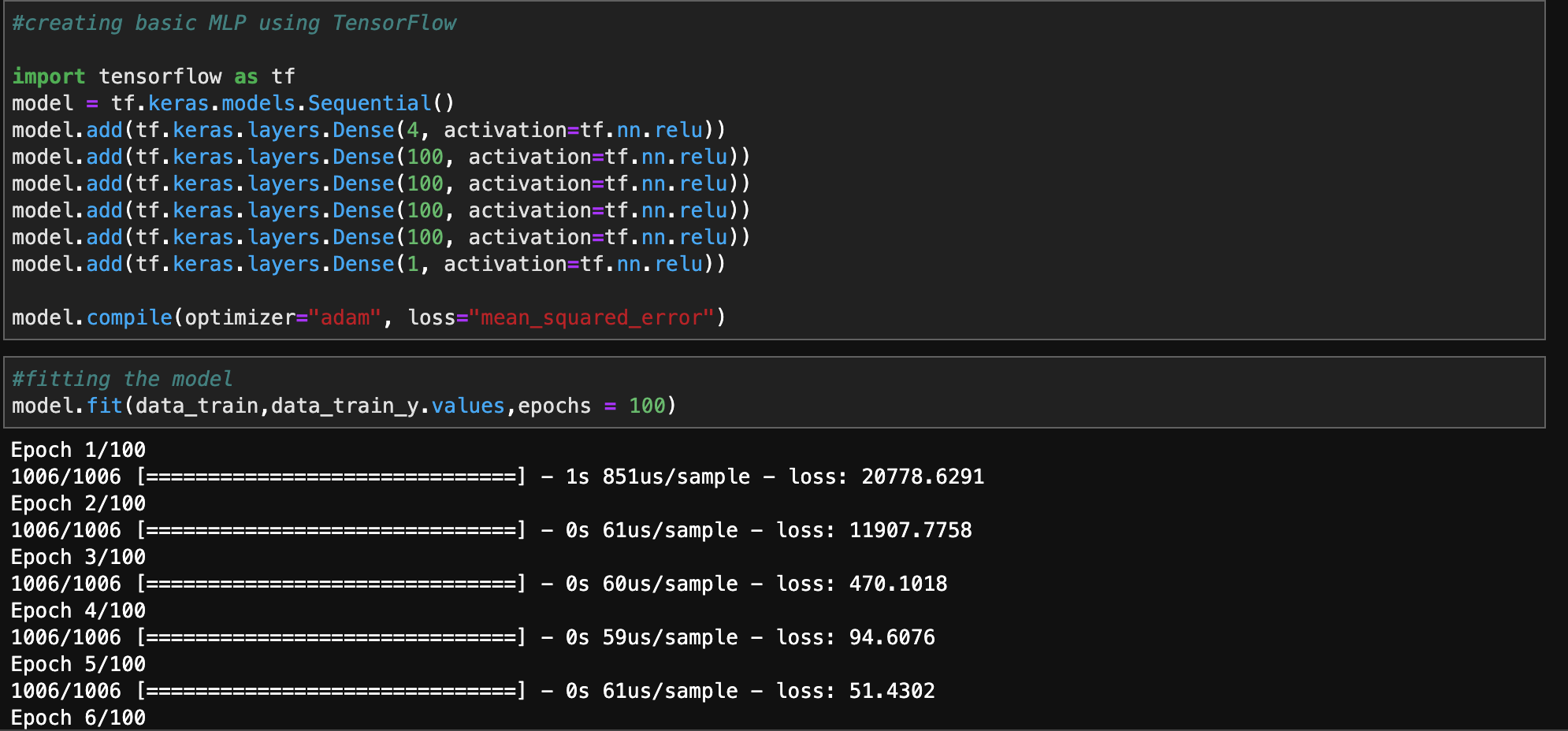
Data Cleaning:



Created a function to clean the data. Normalized the Predictors using Min Max. Removed all rows with any Nan values. Partitioned data for x years as 1:x-1 as train and x-1:x as test.

Neural Network Training:

Created and trained a Neural network as below.



Neural Network Testing:

Evaluated the model based on the Mean Squared Error loss function as follows.



Used Matplot Lib to check if the neural network is able to capture trends:



As we can see above though the neural network is not able to predict exact values it is able to capture trends.

## 4.4 V2 Risk Retirement

Identify and prioritize the 5 top risks in carrying out the project. Try as best you can to retire the top ~~two~~ four by the time you submit this, by means of experiments, prototypes, or work-arounds. Explain how you did this. Explain how you will retire the remaining risks in advance. Update this from the version in Assignment 3.

**Stock Market Uncertainty:**

Stock Market is uncertain since it is affected by factors such as Politics, Natural Calamities, etc.

*Work-arounds:*

This risk is very difficult to mitigate. You will need multiple models such as predicting weather.

**LSTM risks:**

LSTM works in a way that it remembers some of the previous data. You can provide this as an input to the initialization function. This number is usually decided by looking at trends in the data. But if this trend is not followed for some part of the training data the predictions can be affected.

*Work-around:*

We need to try to get as much data as possible and try and set this parameter wisely.

## 4.5 V2 Schedule

Explain in outline the updated steps you intend to take to carry out the project. Show the completion of the stages. Include a schedule, as detailed as can be reasonably foreseen.

Week 4: Implementing basic MLP

Week 5: Implementing LSTM and comparing

Week 6: Trying to implement Nice to have goals

## 4.6 V2 References

Add to your references. Instructions as above.

[1] <https://towardsdatascience.com/the-fall-of-rnn-lstm-2d1594c74ce0>

[2] <https://machinelearningmastery.com/how-to-develop-multilayer-perceptron-models-for-time-series-forecasting/>

[3] <https://blog.rapidapi.com/how-to-use-the-yahoo-finance-api/>

[4] <https://www.geeksforgeeks.org/data-normalization-in-data-mining/>

[5] <https://machinelearningmastery.com/multivariate-time-series-forecasting-lstms-keras/>

[6] <https://www.tensorflow.org/beta/tutorials/text/time_series>

[7] <https://medium.com/@UdacityINDIA/tensorflow-or-pytorch-the-force-is-strong-with-which-one-68226bb7dab4>

## 4.7 Evaluation of Assignment 4

