Project Title: **Microsoft Stock Prices Prediction**

Group Members:

Ibrar Babar(P19-0104)

Noman Siddique (P19-1664)

Abstract:

We trained different deep-learning models to predict Microsoft Stock Prices on Kaggle dataset for better accuracy. And we gain minimum Root Mean Square Error using the LSTM model having RMSE **11.222516%**. Time series forecasting is a very intriguing field to work with. There is a perception in the community that it's a complex field, and while there is a grain of truth in there, it's not so difficult once you get the hang of the basic techniques. I am interested in finding out how LSTM works on a different kind of time series problem and encourage you to try it out on your own as well. If you have any questions, feel free to connect with me in the comments section below.

What exactly is the stock market?

The stock market is a collection of markets and exchanges where regular activities such as buying, selling, and issuance of publicly traded company shares take place. Such financial activities are carried out through institutionalized formal exchanges or over-the-counter (OTC) marketplaces that follow a set of rules. A country or region may have multiple stock trading venues that allow transactions in stocks and other forms of securities.

Understanding the Problem Statement

We'll get into the implementation part of this article soon, but first, we need to define what we're trying to solve. Fundamental analysis and technical analysis are the two broad categories of stock market analysis.

1. Fundamental Analysis is the process of forecasting a company's future profitability based on its current business environment and financial performance.

2. Technical analysis, on the other hand, entails reading charts and analyzing statistical data to identify stock market trends.

As you might expect, we'll concentrate on the technical analysis. To build a model capable of estimating stock prices, we will use the dataset of Microsoft stock prices from April 2015 to April 2021. It's time to get started!

Major Points of Understanding

The dataset contains several variables, including date, open, high, low, close, and volume.

The columns Open and Close represent the opening and closing prices of the stock on a given day.

The maximum and minimum share prices for the day are represented by High and Low.

The number of shares purchased or sold during the day is referred to as volume.

Another thing to keep in mind is that the market is closed on weekends and public holidays.

Take another look at the above table; some date values are missing: 4/3/2015, 4/4/2015, and 4/5/2015. The third of April 2015 was a public holiday due to the occasion of Good Friday, while the fourth and fifth of April were weekends.

Because the closing price of a stock for the day is usually used to calculate profit or loss, we will use the closing price as the target variable. Let's plot the target variable to see how it looks in our data.

Models Used for Prediction Microsoft Stock Prices are explained one by one below.

Moving Average

Introduction

'Average' is one of the most commonly used words in our daily lives. Calculating the average marks to determine overall performance, or determining the average temperature of the previous few days to get an idea of today's temperature - these are all routine tasks that we perform on a regular basis. As a result, this is a good starting point for making predictions on our dataset.

Each day's predicted closing price will be the average of previously observed values. Instead of a simple average, we will employ the moving average technique, which employs the most recent set of values for each prediction. In other words, for each subsequent step, the predicted values are considered while the oldest observed value is removed from the set. Here is a simple diagram that will help you understand this better.

This technique will be applied to our dataset. The first step is to create a data frame with only the Date and Close price columns, and then divide it into train and validation sets to test our predictions.

Observation:

The RMSE value is close to 76.62 but the results are not very promising (as you can gather from the plot). The predicted values are of the same range as the observed values in the train set (there is an increasing trend initially and then a slow decrease).

In the next section, we will look at two commonly used machine learning techniques – Linear Regression and kNN, and see how they perform on our stock market data.

Linear Regression

Introduction

The most basic machine learning algorithm that can be implemented on this data is linear regression. The linear regression model returns an equation that determines the relationship between the independent variables and the dependent variable.

The equation for linear regression can be written as:

Y = mX1 + b

Here, X1, X2,….xn represent the independent variables while the coefficients θ1, θ2, …. θn represent the weights. You can refer to the following article to study linear regression in more detail:

A comprehensive beginner's guide for Linear, Ridge, and Lasso Regression. For our problem statement, we do not have a set of independent variables. We have only the dates instead. Let us use the date column to extract features like – day, month, year, Mon/Fri, etc., and then fit a linear regression model.

Implementation

We will first sort the dataset in ascending order and then create a separate dataset so that any new feature created does not affect the original data.

K-Nearest Neighbours

Introduction

Another interesting ML algorithm that one can use here is kNN (k nearest neighbors). Based on the independent variables, KNN finds the similarity between new data points and old data points. Let me explain this with a simple example.

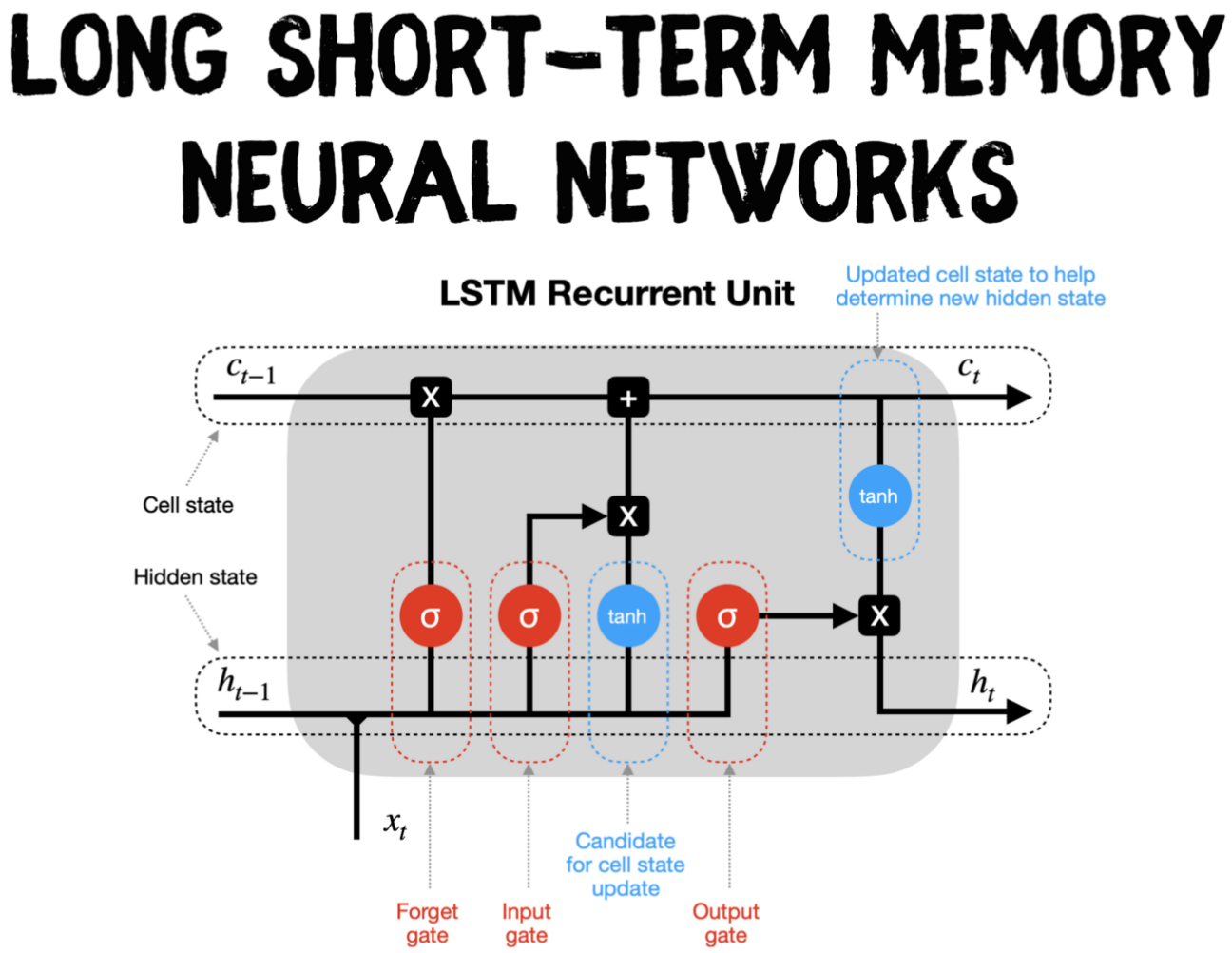
Observation: The RMSE value is almost similar to the linear regression model and the plot shows the same pattern. Like linear regression, kNN also identified a drop in January 2018 since that has been the pattern for the past years. We can safely say that regression algorithms have not performed well on this dataset.

Let’s go ahead and look at some time series forecasting techniques to find out how they perform when faced with this stock price prediction challenge.

Long Short-Term Memory (LSTM)

Introduction

LSTMs are widely used for sequence prediction problems and have proven to be extremely effective. The reason they work so well is that LSTM is able to store past information that is important and forget the information that is not. LSTM has three gates:



The input gate: The input gate adds information to the cell state The forget gate: It removes the information that is no longer required by the model The output gate: Output Gate at LSTM selects the information to be shown as output For a more detailed understanding of LSTM and its architecture,

Observation Wow! The LSTM model can be tuned for various parameters such as changing the number of LSTM layers, adding dropout value, or increasing the number of epochs. But are the predictions from LSTM enough to identify whether the stock price will increase or decrease? Certainly not!

As I mentioned at the start of the article, the stock price is affected by the news about the company and other factors like demonetization or merger/demerger of the companies. There are certain intangible factors as well which can often be impossible to predict beforehand.

Conclusion

Time series forecasting is a very intriguing field to work with. There is a perception in the community that it's a complex field, and while there is a grain of truth in there, it's not so difficult once you get the hang of the basic techniques.

I am interested in finding out how LSTM works on a different kind of time series problem and encourage you to try it out on your own as well. If you have any questions, feel free to connect with me in the comments section below.