

STOCK PRICE PREDICTION

Domain Background

Predicting the stock prices falls under the category of investment and trading, of finance market. Predicting how the stock market will perform is one of the most difficult things to do. There are so many factors involved in the prediction – physical factors vs. psychological, rational and irrational behavior, etc. All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy.

Thousands of companies use software to predict the movement in the stock market in order to aid their investing decisions. Primitive predicting algorithms such as a time-series, linear regression can be done with a time series prediction by leveraging python packages like scikit-learn.

Here is the reference that I found well researching for this Problem

<https://ieeexplore.ieee.org/document/6740438>

A simple TensorFlow model is used here to predict the stock prices.

<https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using-tensorflow-30505541d877>

Problem Statement

The Main Aim of this project is to Predict the stock prices of a particular-company using the stock market dataset. In this project I am going to using use LSTM networks to predict stock prices for a good accuracy. It's important to note that there are always other factors that affect the prices of stocks, such as the political atmosphere and the market. However, we won't focus on those factors for this project.

Datasets and Inputs

The dataset is available in the Kaggle site a reference link is provided below. Which contains all S&P companies' historical data for past 5 years. The dataset contains 619K rows and 7 columns

<https://www.kaggle.com/camnugent/sandp500>

Contents

The main contents of data are

- Date: The date on which we are predicting the stock
- Open: The value at which the stock is opened.
- High: Maximum value of the stock
- Low: Minimum value of the stock
- Close: The value at which the stock is closed
- Volume: The volume of stocks bought
- Name: The name of the company

Solution

We are going to use LSTM networks to predict stock prices. A LSTM network is a kind of recurrent neural network. A recurrent neural network is a neural network that attempts to model time or sequence dependent behavior – such as language, stock prices, electricity demand and so on.

As the data is somewhat small, I am going to train the model upto 300 epochs.

Benchmark

I am going to use Linear Regression as my bench mark model. By using LSTM's, I will try to improve my accuracy.

Evaluation Metrics

I choose my evaluation metric as Mean Square Error (or) Mean Square Deviation. The Mean-square deviation (MSD) or mean square error (MSE) is a frequently used measure of the differences between values predicted by a model or an estimator and the values observed.

Project Design

STEP 1: LOADING THE DATA AND DEFINING THE TARGET VARIABLE

- First the dataset & the required libraries are loaded into the model.
- By analyzing the data, we must choose the target variable, what we are going to predict by constructing this model.

STEP 2: PREPROCESSING THE DATA

- We must pre-process the data, which helps in analyzing the data in a better way.
- MinMaxScaler is used in my model to normalize the target variable.

STEP 3: SPLITTING INTO TRAINING AND TESTING DATA

- We split the data into 7 day look back slices. This data is divided into training and testing data.
- Training data is used to train the model.
- Testing data is used for predicting the target variable

STEP 4: BUILDING THE MODEL

The model is constructed using Keras with TensorFlow as background.

In order to build the LSTM, we need to import a couple of modules from Keras:

1. **Sequential** for initializing the neural network.
2. **Dense** for adding a densely connected neural network layer.
3. **LSTM** for adding the Long Short-Term Memory layer.
4. **Dropout** for adding dropout layers that prevent overfitting.

STEP 5: TRAINING THE MODEL & TESTING THE MODEL

- Using the above Built model with mentioned parameters we train the model.
- The Training is done using the Training data.
- The model is tested on the test data in order to predict the target variable.

SASIKIRAN ADAPA,
MACHINE LEARNING NANO DEGREE,
CAPSTONE PROJECT PROPOSAL,
STOCK PRICE PREDICTION,
WEDNESDAY, February 13, 2019.