

Table of Contents:

1. Overview of Project.................................. 3
2. Introduction.............................................. 4-7

* Motivation
* Aim
* Objectives
* Scope

1. Related Work............................................ 8-12
2. Implementation........................................ 13-20
3. Reference.................................................. 21

Table of Screenshots:

1. Linear Regression Graph…………………7
2. Support Vector Regression (SVR)…….9
3. Implementation (Splitting data) ……..12
4. Linear Regression model………………….14
5. Support Vector Regressor model…… .15
6. Graph of SVR and Linear Regression..16
7. Visualization code snippet……………....17

Overview of Project:

The project which we took up is Stock Market Prediction. In this project we are supposed to predict a company’s stock value for 1 month. Stock Market Prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange.

Stock Market is the important part of economy of the country and plays a vital role in the growth of the industry and commerce of the country that eventually affects the economy of the country. Both investors and industry are involved in stock market and wants to know whether some stocks will rise or fall over a certain period.

The stock market is the primary source of any company to raise funds for business expansion. It is based on the concept of demand and supply. If the demand for a company’s stock is higher, then the company share price increases and if the demand for company’s stock is low then the company share price decreases.

Introduction**:**

* Motivation:

The motivation for the project in this field is that it possesses many theoretical and experimental challenges.

The most important of these is the Efficient Market Hypothesis (EMH), the hypothesis says that in an efficient market, stock market prices fully reflect available information about the market and its constituents and thus any opportunity of earning excess profit ceases to exist.

One of the examples of big stock exchange is New York Stock Exchange (NYSE)

* Aim**:** The aim of this project are as follows:

To identify factors affecting share market

To generate the pattern from large set of data of stock market for prediction of NSEM

To predict an approximate value of share price of a company’s stock prices.

* Objectives:

In the past decades, there is an increasing interest in predicting markets among economists, policymakers, academics and market makers. The objective of the proposed work is to study and improve the supervised learning algorithms to predict the stock price.

Technical Objective:

The technical objectives will be implemented in Python. The system must be able to access a list of historical prices. It must calculate the estimated price of stock based on the historical data. It must also provide an instantaneous visualization of the market index.

Experimental Objective:

Two versions of prediction system will be implemented; one using Linear Regression and other using Support Vector Machines. The experimental objective will be to compare the forecasting ability of SVM using Support Vector Regression (SVR). We will test and evaluate both the systems with same test data to find their prediction accuracy.

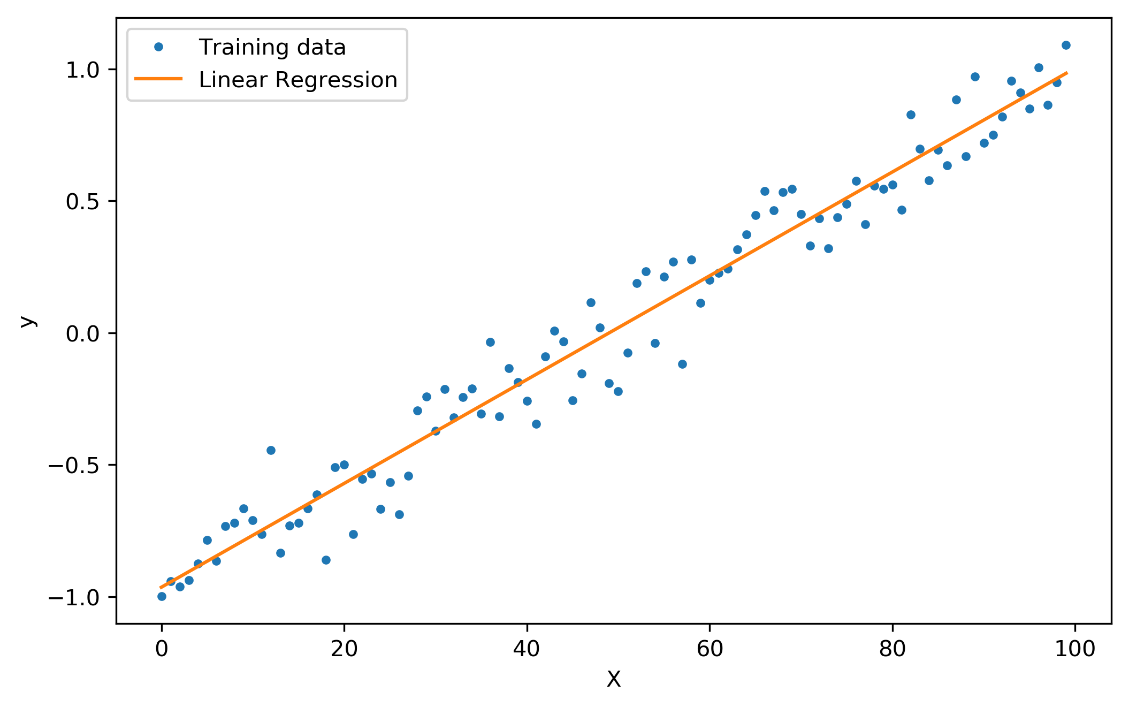
* Scope:
  + The scope of this project does not exceed more than a generalized suggestion tool.
  + Analysis of stocks using data mining will be useful for new investors to invest in stock market based on the various factors considered by the software.  
    Stock market includes daily activities like Sensex calculation, exchange of shares.
  + This project requires investigation in the following areas: Stock Market [1] Investigating trends in stock market and factors affecting the stock prices.
  + Data mining techniques Investigating the available tools and techniques for data mining and then selecting those that are best fit to solve the problem.

Related Work:

About Linear Regression:

What is linear regression?

When we see a relationship in a scatterplot, we can use a line to summarize the relationship in the data. We can also use that line to make predictions in the data. This process is called linear regression.



* + Here the training data represents the data we’re going to use.
  + Linear Regression is the data predicted or extracted after training the data using Linear Regression.

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.

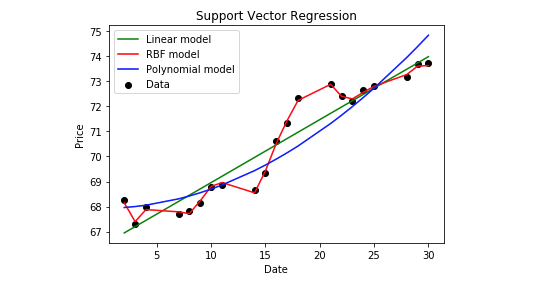
Linear Regression is an inbuilt function/class from the library sklearn/sklearn.linear\_model in python.

Before attempting to fit a linear model to observed data, a modeler should first determine whether there is a relationship between the variables of interest. This does not necessarily imply that one variable *causes* the other (for example, higher SAT scores do not *cause* higher college grades), but that there is some significant association between the two variables.

A scatterplot can be a helpful tool in determining the strength of the relationship between two variables. If there appears to be no association between the proposed explanatory and dependent variables (i.e., the scatterplot does not indicate any increasing or decreasing trends), then fitting a linear regression model to the data probably will not provide a useful model.

About Support Vector Regression:

As in classification, support vector regression (SVR) is characterized using kernels, sparse solution, and VC control of the margin and the number of support vectors. Although less popular than SVM, SVR has been proven to be an effective tool in real-value function estimation.



* Here Support Vector Regression (SVR) has 3 various types of model which has set of attributes to gives the expected output.
* Among those 3 models RBF model gives the more accurate prediction for us, it aligns the data-values like shown in the above graph.

Support Vector Machine can also be used as a regression method, maintaining all the main features that characterize the algorithm (maximal margin). The Support Vector Regression (SVR) uses the same principles as the SVM for classification, with only a few minor differences.

First, because output is a real number it becomes very difficult to predict the information at hand, which has infinite possibilities. In the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM which would have already requested from the problem.

But besides this fact, there is also a more complicated reason, the algorithm is more complicated therefore to be taken in consideration.

The benefits of using SVM over neural networks are:

1. SVM has strong founding theory.
2. Global optimum guaranteed.
3. Requires less memory to store the predictive model.
4. Yield more readable results and a geometrical interpretation.

Why not Neural Networks:

The other approach makes use of Neural Networks.

Neural networks have the following drawbacks:

1. Slow Convergence Rate:

The neural network takes a lot of time to train.

2. Local Minima and Maxima:

Neural Networks are based on gradient descent method to find the local extreme value and they tend to get stuck on the local minima and maxima.

So, therefore it is difficult to find global minima and maxima.

Implementation:

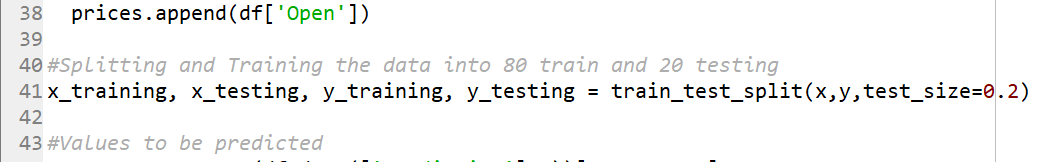
Data Input:

Input data is taken by following steps:

1. For our project, we are considering Google’s stock prices over past few years (for better results).
2. We will import pandas library to get access and modify the data that we imported.

3. By using pandas library we’ll induce the dataset and modify what we want to make use of it.

4. Further we divide the data into two parts, training data and testing data, where 80% of the data will be used for training and 20% of the data will be used for testing.



For training and testing we need to divide the data we do this by importing the train\_test\_split class or function.

How to get the Prediction?

* Linear Regression
* Support Vector Regression

To solve the problem, we will follow below steps:

1. Fetch the data of a stock from Google’s stock prices list from Yahoo Finance.

2. Visualizing the data collected datasets as per requirements of the outcome.

3. Train the model using the Predictive models and training data.

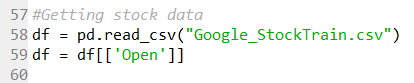
4. Test the model using testing data.

5. Evaluate our system using various evaluation techniques.

How to Generate the Output?

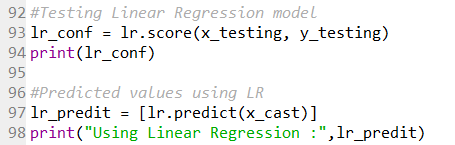
Using Linear Regression:

1. Get the data.



b) Provide the data to the system.

c) Train the system.



d) Model will predict the output.

The output is also included in the above screenshot.

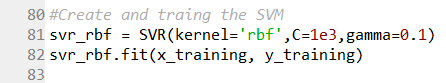
Using Support Vector Machines:

We will be using C-classification Support Vector Machine with RBF Kernel.

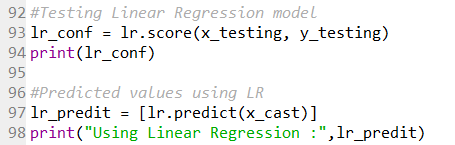
Step 1: Read the required data [Date, Open, High, Low, Close, Volume, Adjusted]

Step 2: Calculate the prediction variable (Up/Down)

Step 3: Provide the data from above steps to train SVM (RBF, C = 1, gamma = ½)

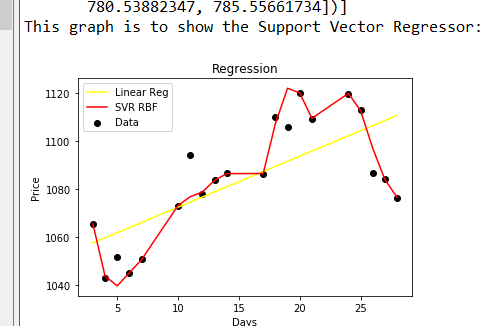


Step 4: Provide test data and display the results

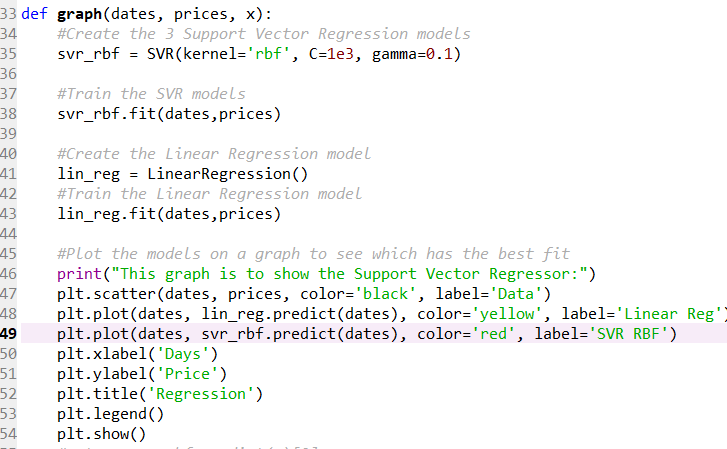


What’s new?

* Visualization plays an important role while dealing with datasets.
* For the datasets and the predicted values, we used scatter plot to represent the data in graphical way.



* By using matplotlib library plotting the graph has made much simpler.
* The graph is between the prices and days from the datasets and the predicted values.
* Method used for the graph part is given below:



* Here the required data is being taken for visualizing the scatter graph and displaying.

Conclusion:

SVM is a promising type of tool for financial forecasting. SVM is superior to the other individual classification methods in forecasting daily movement direction. This is a clear message for financial forecasters and traders, which can lead to a capital gain.

However, each method has its own strengths and weaknesses. In this model, the principal components identified by the SVM are used along with internal and external financial factors in SVM for forecasting.

We also observed that the choice of the indicator function can dramatically improve/reduce the accuracy of the prediction system.

References:

* https://www.statisticssolutions.com/what-is-linear-regression/
* https://www.saedsayad.com/support\_vector\_machine\_reg.htm
* https://in.finance.yahoo.com/
* https://support.google.com/docs/answer/3093281?hl=en
* https://www.khanacademy.org/math/statistics-probability/linear-regression-review