Online learning of Gaussian Process

Stock Price prediction using Gaussian Process

Hrithik Singh (20857)
Department of
Management Studies,
IISc
hrithiksingh@iisc.ac.in

Abstract

Historical stock price data is a huge amount of time series data with negligible noise. With the help of such data ideally we should be able to predict future stock prices with great accuracy. It turns out to be a challenging task because of the high volatility in the stock market values. It is evident that there are many factors such as physical , political and psychological that affect the stock price of a company. To investors stock market is of key importance because it involves large amount of frequent money transactions. In an attempt to predict the future stock prices we use gaussian process regression.

1 Introduction

Stock market is basically defined as collection of exchanges where share of different companies can be bought, sold or issued. Hence stock market provides a common platform for trading stocks of different companies in real time either in physical or in financial format.

Share prices of different companies on stock exchange can be determined in many different ways. One of the most common method is to auction it. Buyers bid on the share and companies offer the price at . which they are willing to sell. A trade is executed when the stock's value is listed, and the buyer matches their quoted price. Millions of investors and traders combinedly form the stock market. On a normal day, millions of dollars’ worth of transactions happen between an investor and trader. Market conditions influence stock value every day. Thus, stock prices fluctuate due to fluctuations in supply and demand. When the demand is more than the supply, the price rises. On the other hand, if the supply exceeds the demand, the price falls.

The worth of a company depends on how it is generating its revenue. The earnings of public corporations are reported four times a year (once in each quarter). The stock analysts at big stock exchange firms use the earnings forecasts to determine a company's future value. If a company's results are on a positive note, the share price of that company increases while if a company's results are on the negative side then the share price of that company falls.

2 Literature Survey

Most of the improvements that have been seen in recent times have come from the field of Machine Learning. A decade ago, predicting stock price was a time consuming and tedious process, but now this has become easier by using machine learning and deep learning algorithms for the stock market projections. These techniques have outperformed humans' ability to predict. Machine learning techniques are based only on facts, figures, and data, without considering any emotions and other psychological factors. Some of the methods which are being used for stock price prediction are discussed below.

Linear regression is one of the methods which can be used to predict the stock price. But it has certain drawbacks when applied throughout a larger time horizon. Linearity might not be that prominent. For some companies , there might be a very small correlation because of price saturation. There are problems even if we apply it on smally time horizon. There could be a case of overfitting.

Auto-Regressive Integrated Moving Average (ARIMA) is also very popular method for forecasting of time series data. In ARIMA, there are three key parameters: p (values of past that are used for future predictions), q (error that have been made in past) and d (order of differencing). ARIMA model has many hyper-parameters that require careful tuning. The Auto AROMA model is hence more preferred to reduce the overhead of careful tuning, which routinely finds the best values for (p,q,d) with the least amount of error. This model is able to predict an upward or a downward trend in the prices using these variables.

2.1 Gaussian Process

It is stochastic process where the random variables (which can be indexed with time or space) have multi-variate normal as their joint distribution and hence their individual marginal distribution is also normal. Every gaussian process curve specifies a single observation form the multi variate joint distribution. To state this even more clearly we can say that if we specify two random variables at different point of time and make scatter plot of points (values of different curves for these two random variable) then we will get a bi variate normal distribution.

As we know that random variables can be finite and can also be infinite in numbers in a given stochastic process. Hence in order to completely define a gaussian process we specify mean function and covariance function. The mathematical formulation goes as follows :

Let x be some process (x).

$$f(x) \sim GP(m(.), k(.,.)),$$

where m(.) denotes mean, and $k(.,.)$ denotes covariance function, respectively.

$$m(x) = E[f(x)]$$

$$k(x_1.x_2) = E[f(x_1) - m(x_1))(f(x_2) - m(x_2))]$$

For a training set, $D = \{(x\mathrm{i}, Y\mathrm{i}) \mid \mathrm{i} = 1, 2, \dots, N\}$, where $x_\mathrm{i} \in R^D$ is the input vector and $y_i \in R$ is the vector of target variables. The observations y_i from the process *f(x)* are assumed to be noisy:

$$y_\mathrm{i} = f(x_\mathrm{i}) + \mathrm{s}_\mathrm{i}, where\ \mathrm{s}_\mathrm{i} \sim N(0, \sigma^2)$$

For simplicity, the GP prior distribution is assumed to have mean zero $f(x) \sim GP(0, k(.,.))$.
Let $f = [f(x_1), \dots, f(x_n)]$ be a vector of function values in the training set D. Their prior distribution is as follows:

$$f \sim GP(0,\ K(\mathrm{X},\ \mathrm{X})),$$

where $\mathrm{K(X, X)_{ij}} = k(x_\mathrm{i}, x_\mathrm{j})$ is a covariance matrix evaluated using covariance function between given points (also known as kernel or Gram matrix).

When it comes to prediction, we predict the expected value of the marginal distribution(normal) of that random variable.

$$E[f_*|y] = K(\mathrm{X}_*, \mathrm{X})\, \Lambda^{-1} y,$$
$$Cov[f\ |y] = K(\mathrm{X}\ , \mathrm{X}\ ) - K(\mathrm{X}\ , \mathrm{X})\Lambda^{-1} K(\mathrm{X}, \mathrm{X}\ )$$

where y are test points and

$$\Lambda = K(\mathrm{X}, \mathrm{X}) + \sigma^2 I_N \qquad n$$

The computation of Λ^{-1} as the heaviest in terms of requirement of computational power in Gaussian process regression, with time and space complexity as $O(N3)$ time and $O(N2)$ respectively.

For stock price data we try different individual kernel functions and combinations of different kernel functions for getting best prediction on future prices of a stock. One more thing we do is to normalize the avg stock price to zero, because that is the prior we initially proceed with.

3 Methodology

- Dataset- Daily and monthly stock price data of Apple Inc. will be used dating from 01-01-2013 to 31-03-2023. Data is downloaded in the form of a csv file from the website of Yahoo Finance.
- Applied Gaussian process to predict the average future monthly stock price using different sizes of training dataset. Also predicted past missing values of average monthly stock prices.
- In order to get prediction with less variance we normalized the stock prices to zero mean because in gaussian process mean function is assumed to be zero.
- Future stock prices are affected by trends and patterns, which are up to 2 years old. Generally, more the data more will be the confidence in prediction. Now if we apply gaussian process on daily data then total data points will be around 700 which will lead to a very high time complexity for the algorithm. If we take only one year data, then it will not give a sound prediction. Now I order to capture the relevant time span and to keep the data points less, we applied the gaussian process algorithm on average monthly stock price data. Here total data points were 24 in the training set.
- Finally, we took two years data to predict the stock price of every month from 01-01-2018 to 31-03-2023 and calculated the mean absolute deviation of the predicted values.
- To do this we analyzed different kernel combinations by plotting their graphs. On y-axis we had covariance value and on x-axis we had value of x2, x1 was fixed to zero.
- We explored various kernels like RBF, matern (1.5), matern (2.5), exponential sine squared and constant kernel for making different combinations.
- We also did hyper-parameter optimization on different kernel combinations to get minimum possible mean absolute deviation.

4 Conclusion

- Code of gaussian process and different kernels from scratch.
- The following kernel combinations were analyzed and kernel_8 gave the minimum mean absolute deviation of **5.005.**
  - (a) Mt – Matern kernel
  - (b) RBF – Radial basis function kernel
  - (c) C – Constant kernel
  - (d) ExpSS – Exponential sine squared kernel

[6]:
```
def kernel_1(x,y):
    return C(12.8**2)*Mt(x,y,length_scale=2.08, nu=1.5)+ExpSS(x,y,length_scale=3.08, periodicity=6)
```

[7]:
```
def kernel_2(x,y):
    return C(12.8**2)*Mt(x,y,length_scale=2.08, nu=2.5)
```

[8]:
```
def kernel_3(x,y):
    return C(12.8**2)*RBF(x,y,length_scale=20)
```

[9]:
```
def kernel_4(x,y):
    return C(12.8**2)*ExpSS(x,y,length_scale=2.08, periodicity=6)+C(12.8**2)*RBF(x,y,length_scale=5)
```

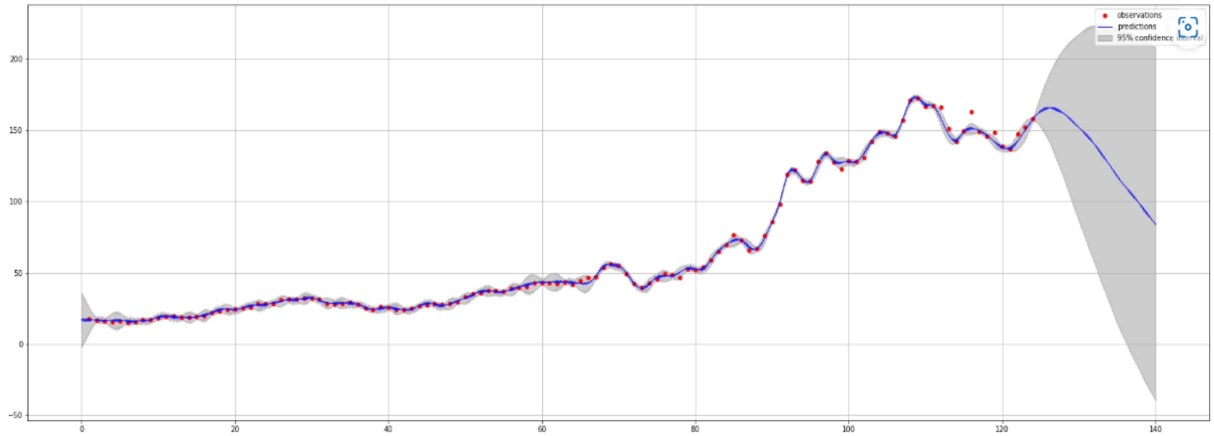
[10]:
```
def kernel_5(x,y):
    return C(12.8**2)*ExpSS(x,y,length_scale=2.08, periodicity=6)+C(5**2)*RBF(x,y,length_scale=5)
```

[11]:
```
def kernel_6(x,y):
    return C(12.8**2)*ExpSS(x,y,length_scale=2.08, periodicity=6)+RBF(x,y,length_scale=5)
```

[12]:
```
def kernel_7(x,y):
    return C(12.8**2)*Mt(x,y,length_scale=2.08, nu=1.5)+RBF(x,y,length_scale=5)
```

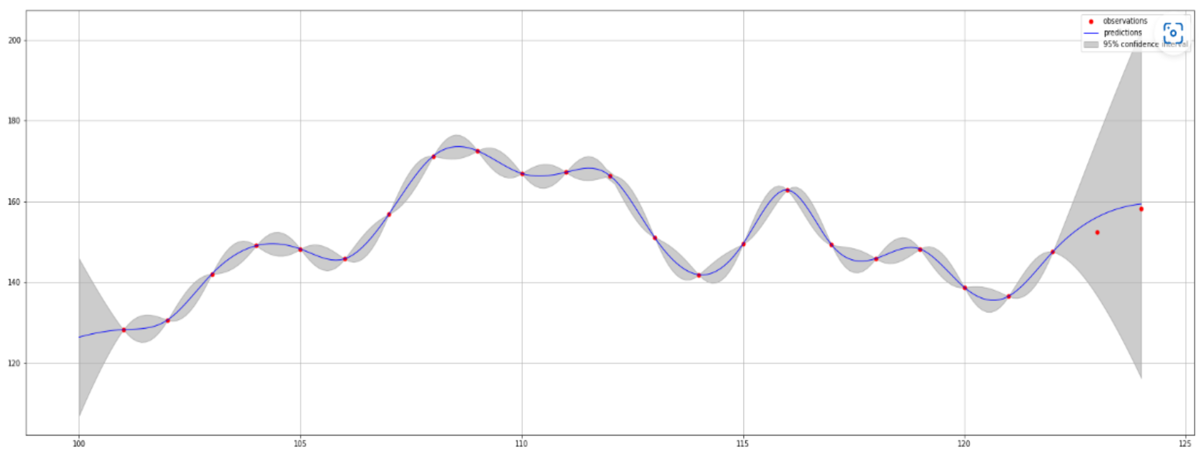
[13]:
```
def kernel_8(x,y):
    return Mt(x,y,length_scale=2.08, nu=2.5)+C(12.8**2)*RBF(x,y,length_scale=5)
```

- Below is the snapshot of prediction and the actual value of the avg stock price for past missing data in the span of past 10 years:



From the above figure it is clear that by using Gaussian Process, we can predict past missing stock price datapoints very accurately with very high confidence.

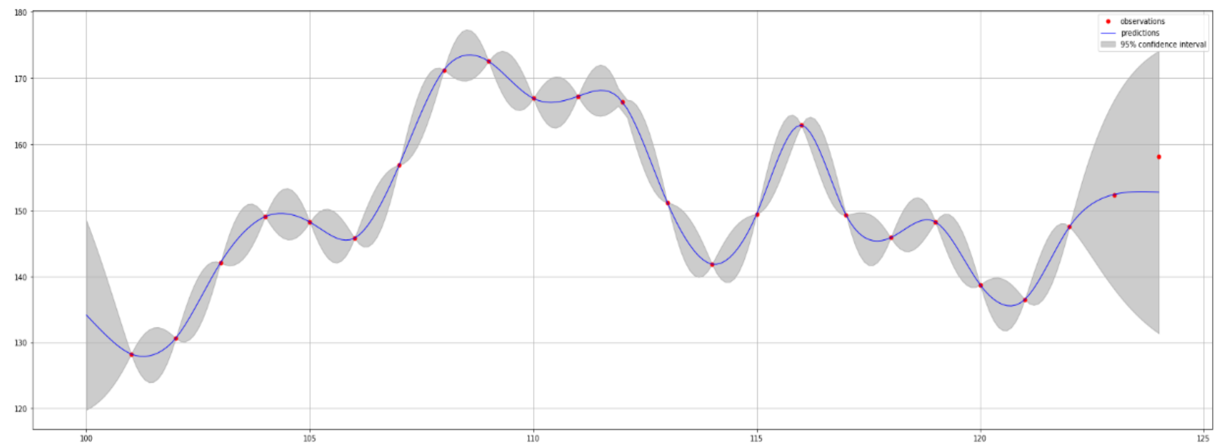
- Below is the snapshot of prediction and the actual value of the avg stock price for the month of February and March without normalization :




```
Mean absolute deviation = 2.4946012553080124
```

Here we got a pretty accurate prediction but the 95% confidence interval is very wide (118-210) because this graph was made without normalizing the stock prices.

- Below is the snapshot of prediction and the actual value of the avg stock price for the month of February and March with normalization :




```
Mean absolute deviation = 2.718141248504878
```

Here the accuracy of the prediction is slightly lesser than above case but the 95% confidence interval is a lot more sharper (132-175).

5 Questions

- What are the differences between parametric and non-parametric methods?

  In parametric method we assume a particular type of function (hence the number of parameter is fixed) for modelling the input and output in advance, and we learn the parameters of the assumed function from the data in order to minimize the loss function. E.g. Linear Regression

  In non-parametric methods, we do not make any assumptions about the function for modelling the input and output in advance. Hence number of parameters is not fixed. These method adapt to the complexity of data to estimate the function. E.g. Gaussian Process. Here the model is defined by some mean and covariance function but their parameters are not fixed in advance, instead they are learnt form the data itself.

- What are the challenges in using gaussian process in an online setting?

  Following are the challenges which I faced in this project:

  (a) Since the data were coming sequentially and we used past 2 year data to predict the next stock price value in future, we had to optimize the parameters of covariance function for each and every prediction because of the change in training data which was time consuming.

  (b) Since we were doing optimization again and again, predictions were very good for most of the test points but sometimes predictions were way of because of limited training data.

- How to choose the hyper parameters?

  Following are the factors which were considered for choosing hyper parameters in this project:

  a) Is there any trend in data.

  b) Is there any periodicity.

  c) How far away are the consecutive data points in the sequence.

  For optimization, we used telescopic search. First we computed the loss functions by taking different values of hyper parameters separated by a fixed gap. Then we zoomed in the region where we were getting minimum mean absolute deviation.

- How to model non-stationary data using GP in an online setting?

  For doing this we used exponential sine squared kernel. Now since for every prediction our training data was different (past 2 years) and we were optimizing the parameters for each prediction, hence we got different parameter values for different time stamps.

- Time complexity and scalability analysis of Gaussian Process?

  Time complexity of a GP model can be expressed as $O(n^3)$, where n is the number of datapoints in training data set. This is because the most computationally expensive operation in the GP models is inversion of (n x n) kernel matrix, which has a computation cost of $O(n^3)$. Hence as the number of training datapoints increase the computation cost of models grows rapidly.

  We faced this problem when we tried to predict the future data points by training on the entire previous data of 10 years. In order to cover a large time span and also to have less datapoints we finally analyzed the avg monthly stock price by training the model on past two years data (24 datapoints) . This gave us reasonable span to cover the previous trends and patterns and less data points to run the GP model efficiently.

  By doing this we also got better results because anyways for a volatile dataset like stock market price, way older trends have no effect on the current trend.

6 References

[1] Proceedings of the Sixth International Conference on Computing Methodologies and Communication (ICCMC 2022) IEEE Xplore Part Number: CFP22K25-ART; ISBN: 978-1-6654-1028-1
https://ieeexplore.ieee.org/document/9754114

[2] M. Todd Farrell, Andrew Correa. Gaussian Process regression Models for predicting Stock Trends.
https://www.researchgate.net/publication/249769503_Gaussian_Process_Regression_Models_for_Predicting_Stock_Trends