

University of Electronic Science and Technology of China

Thesis Proposal Form of Grade 2017 of Bachelor's Academic Degree

Student ID:	Name:	School: Information and	Major:
2017221501010	Etim Ubong-Abasi Ebong	Software Engineering	Software Engineering
Thesis Title	Gaussian Process Prediction of Stock Price Trends.		
Research Source (Choose one): 1			
1. Scientific Research 2. Production 3. Teaching (including Experiment) 4. Innovation and Entrepreneurship			
Form of Expected Results (Choose one): 4			
1. Hardware 2. Hardware & Software 3. Software 4. Theoretical Research			
Research Content	<p>1. Research status and development trend</p> <p>In the last couple of years, the financial industry has been greatly affected and positively so by the emergence of digitalization and machine learning. However, a few methods that have undergone rapid development in recent years include: Gaussian processes and Bayesian optimization. The world of finance has become more volatile than it has ever been hence, it has been only necessary that there be tested and methods with which losses or even drastic changes in general (profits as well as losses) in the stock market for example can be predicted in order to better prepare companies and firms for them and maybe even arm these companies with the tools proper with which they can properly mitigate these threats to their income. The advent of these machine learning methods and has been extremely crucial to this. There are numerous gaussian methods that have become without a doubt helpful when it comes to finance and stock market predictions and as a result of ongoing research in the fields of machine learning and finance, these methods are only going to become more prominent and numerous in the years to come. Some of the gaussian methods of research that have been employed so far include Gaussian Process Regression, Twin Gaussian Process and many more. This project explores the secrets behind conducting successful predictions of stock market indices.</p> <p>2. The basis and significance of the topic</p> <p>The main focus of this project is to acquire knowledge of the mathematical conceptualizations involved in Gaussian Processes and implement them on in real-world problems The project involves taking prerecorded official stock prices and training them using Bayesian inference to derive a gaussian process regression model, marginal likelihood, etc. to be able to make accurate (sometimes inaccurate) predictions.</p> <p>Over time, the model can be used to derive other predictions for the stock price data for other companies and organizations. The project was completed in python programming language using SKLEARN and MATPLOTLIB libraries mainly because python programming language is easy to use and is also known</p>		

for its wide verity of uses especially in the fields of data science and machine learning.

3. Research contents

The purpose of this thesis is to point out or give a preamble of the remaining content of this paper.

In Chapter 2 will be the fundamental introduction of the theory and method of Gaussian process. Section 2.1 gives a basic idea of exactly what a “Gaussian Process Regression Model” should look like. Section 2.2 expatiates on the GP prior and the formulae involved. 2.4 gives the formula for the marginal likelihood. 2.3 is about the Hyperparameters learning or model optimization. And 2.5 is the formula for making predictions.

Chapter 3 is all about the project decomposition, basically an explanation of what the project is and what the project should do including the method/methodology of forecasting.

Chapter 4 is the implementation of the Gaussian process method on stock market. Section 4.1 expatiates on the dataset process and the stock data to be used for the experiment. In section 4.2, we see the graphical representations of the training (model fitting) and testing (graphical representation of predictions) processes. In section 4.3, we get to see some of the source code used for some of the processes in this project.

4. Major problems to be addressed and ultimate goals; main theories, technical routes and implementation plans to be adopted, etc.

Major problems to be addressed and ultimate goals: making the changes in stock market price trends to become more predictable by gaining knowledge about the mathematics involved in solving problems using gaussian (and by extension, Bayesian) inference for the purpose of applying this new found knowledge to solve real world stock market problems and also aid investors with tools to be able to make mathematically informed decisions about their investment choices.

Main theories: main theories include understanding of the gaussian process, gaussian process regression and its components like the gaussian process model, prior and posterior distributions, marginal likelihood, dataset preparation processes, model fitting techniques and prediction.

Technical routes and implementation plans to be adopted:

These include implementing the mathematical understanding of gaussian process regression using programming (in python programming language) to implement the mathematical theories, configuring the dataset to be used in order to ensure that it is ready for use to get the best possible prediction outcomes.

This thesis implementation plan is as follows:

(1) Algorithm design

The algorithm design entails the step-by-step processes involved in transforming AAPL data and making sure that it became usable for deriving the

	<p>somewhat accurate predictions. It also involves</p> <p>(2) Dataset preprocess</p> <p>The dataset preprocess contains the formatting information, the details of the cleaning process and the hyperparameters used.</p> <p>Price trend observations: The upward or downward trends in the prices of AAPL stocks are observed and then a graph of normalized price trends is created from the normalized prices (normalized: not necessarily affected by splits or dividends).</p> <p>The observations of the normalized price trends: The algorithm is used to observe the upward or downward trends in adjusted and normalized closed prices. The idea is to forecast or predict the trend in prices as accurately as possible against the actual observations taken during the training period.</p> <p>The price trend predictions: The trend of the predictions is plotted against the actual price trend observations in order to see the level of accuracy that the algorithm is capable of.</p> <p>5. Thesis characteristics or innovation points</p> <p>What's new: this project focuses on some new and innovative ways to perform gaussian process regression. One instance where this is visible is in the hyperparameters. Usually, the problem of hyperparameters is solved by minimizing validation errors through the process of cross validation which tests a model's prediction ability based on training data but in this project however, the process of hyperparameter optimization is undertaken by the method of maximizing the marginal likelihood or maximizing the log likelihood in order to obtain the desired hyperparameters. Secondly, the cleaning process of the AAPL dataset is one which is divided into two phases which are equally important in ensuring that the dataset is free of redundancies and repetitions and any other factors that might limit the ability of the time series to be used for prediction from predicting the price trend as accurately as possible. The first phase of this cleaning process involves thorough checking of the dataset by importing the .csv file to a jupyter notebook where a series of commands are run in order to make sure that the dataset does not contain any redundant or empty data points. The second phase of the dataset cleaning is done to ensure an average of 252 trading days in a year and that the price scale starts at zero for all prices and does not exceed a total of 4 (units of the scale) in order to give us proper context of the prediction.</p> <p>How is it beneficial: this approach is beneficial for two main reasons. One, it allows for exploration of a different way of optimizing hyperparameters and fitting model parameters using python's sklearn library optimizer "fmin_l_bfgs_b".</p>
Advice from	<p>Further research should be paid on the method and the algorithm of the Gaussian process regression, especially how the model is trained (or how the hyperparameters are learned from observations).</p>

Supervisor	Signature: <i>Weidong Wang</i> Date: 2021.04.21
-------------------	--