ISSN 2616-5805 Vol. 5, Issue 3: 44-51, DOI: 10.25236/AJMS.2024.050307

Gaussian Process Regression Model Based on Multi-Kernel Learning and Bagging Algorithm

Jiaru Wang[1], Zhenhan Chen[2]

[1]*School of Applied Mathematics, Shanxi University of Finance and Economics, Taiyuan, 030000, China*
[2]*School of Economics, Xinjiang University of Finance and Economics, Urumqi, 830000, China*

Abstract: *The Gaussian process regression model is a flexible and powerful non -parameter regression method. It can fit the relatively complicated function of the Reproducing Kernel Hilbert Space (RKHS) by selecting the appropriate kernel function. However, the correctness of the selection of the kernel function has an important effect on the efficiency of the predictive model and there is no uniform paradigm for solution. Based on this, this article proposes a new self-adaptive Gaussian process regression model by using multiple kernel function. On the one hand, this proposed method can fit the predicted function in a more large RKHS, and adapts to solve the selection problem of kernel function. On the other hand, this method enhances the stability of the Gaussian process regression model by using the Bagging algorithm. In addition, the basis model in the Bagging framework is model-free. The simulation data analysis shows that the proposed model has smaller mean errors, relative errors and absolute errors relative to some comparison methods. Lastly, it is applied to the actual stock data of quantitative transactions and the results also reflect the accuracy and stability of the proposed model.*

Keywords: *Stock Price Modeling, Gaussian Process Regression, Multi-Kernel, Bagging Algorithm*

1. Introduction

In the financial problems, the relationship between the stock price factor (X) is complex and the relationship between these factors and the stock price (Y) is uncertain. When the relationship between the factor and the stock price is linear, researchers often use a penalty linear regression model. Such as, Li Bochao based on the ridge regression method, the specific factors that affect stock prices are studied and the ridge regression parameter k variation chart is drawn to screen out the independent variables that have a practical impact on the prediction results. Secondly, the second-order multiple linear regression cross-term is introduced to construct the model and statistical methods are used to test the significance of the prediction results and the actual results [1]. Wang Guo, Liang Bake Ting and Wang Jinzhi proposed an adaptive Lasso model [2]. Shu Shike and Li Lu put the cross-entropy loss function of logical regression replaced in the elastic mesh. The average error loss function was established, thereby establishing a new type of logical regression to elastic mesh model (LR-RasticNet), which applies the model to the stock price prediction [3]. However, in the stock price forecast, the relationship between the factor and the stock price may be non-linear. In this case, nonparameter regression methods has become powerful tools. For example, Yan Zhengxu, Qin Chao and Song Gang proposed a new combination model that combines the Pearson coefficient based on the stock price prediction method of random forests [4]. Tian Liwei has studied the issue of stock selection from a classification perspective, and proposed a multi-layered LSTM-BO-Lightgbm model in stock fluctuation prediction [5]. In addition to the above models, the decision trees and K-neighbors regression are also common methods for processing non-linear relationships. For instance, the random forest algorithm proposed by Breiman (2001) improves the robustness and prediction accuracy of the model by building multiple decision trees and integrating their prediction results [6]. In the realm of stock price forecasting, the integration of future income uncertainty bears profound ramifications for investment decision-making. More precisely, the inclusion of uncertainty quantification measures, such as interval forecasts delineating potential upward and downward movements,augments investors' capacity to ascertain optimal investment allocations and deploy risk mitigation strategies grounded in these supplementary insights. In this context, Gauss's process regression (GPR) emerges with distinctive advantages. GPR not only furnishes predictive estimates but also furnishes associated uncertainty assessments, thereby endowing it with substantial utility in addressing the quantitative challenges inherent in stock price transactions. For instance, Yang Zhenli and Xia Kewen introduced a Gaussian process regression

ISSN 2616-5805 Vol. 5, Issue 3: 44-51, DOI: 10.25236/AJMS.2024.050307

method utilizing the particle group algorithm to forecast the initial public offering prices of new shares. Consequently, the model not only enhances the accuracy of stock price predictions but also elucidates the associated prediction uncertainties [7-8]. Additionally, Zhu Hongyu employed a cross-validation technique to assess the efficacy of the Gaussian process model in predicting stock prices, affirming its feasibility [9].

However, the selection of an optimal kernel function poses a significant challenge in Gaussian Process Regression (GPR). The choice of kernel functions critically impacts the model's performance and predictive accuracy, with different kernels exhibiting varying degrees of suitability for specific datasets. In practical applications, researchers often tailor their selection of kernel functions according to the underlying characteristics of the data [10-11]. For instance, datasets exhibiting periodic behavior may benefit from a periodic kernel function, while those with higher degrees of smoothness may be better suited to a square exponential or radial basis function (RBF) kernel. Moreover, researchers take into account the dimensionality and scale of the data when considering the computational efficiency of candidate kernel functions. Despite the availability of multiple kernel functions, there exists no universal standard or paradigm to guide the selection process,making it a nuanced and context-dependent endeavor.

To tackle this issue, this article proposes an adaptive Gaussian process regression model utilizing multiple kernel functions [12-14]. Additionally, to enhance model stability, this article introduces the Bagging algorithm, aggregating predictions from multiple base models through weighted averaging [15-16]. To validate our methods, this article conducts both synthetic data analysis and practical application research. In the synthetic data analysis, this article compares our approach with common regression methods, using metrics such as mean square error, relative error and absolute error. Through diverse evaluation metrics, this article demonstrates the superiority and stability of our method in predicting complex functions.

2. Theory and method

2.1 Gauss Process Regression Model

The principle of the Gaussian process regression model is that by introducing the Gaussian process as a priority distribution of data, the correlation between the kernel function is used to describe the correlation between data, thereby building a flexible regression model. Under the Bayesian framework. First of all, this article defines a set containing different kernel functions and allocate a unique serial number for each kernel function. These kernel functions can be a linear kernel function, a polynomial nucleus, the Gauss kernel function (RBF) or the Sigmoid kernel function. Next, let's briefly introduce the principle of the regression of Gauss's process and the common kernel function form. The common kernel function formulas are specific:

① Linear kernel function (kernel1):

$$K(x, y) = x \times y \tag{1}$$

Features: simple and intuitive, suitable for linearly divided.

② The polynomial kernel function (kernel2):

$$K(x, y) = (\gamma \times x \times y + c)^d \tag{2}$$

Features: It can handle non-linear relationships, but it is necessary to pay attention to choosing the right parameterγ, c and d to avoid overfitting or arrears.

③ Gaussian kernel function (radial base function RBF) (kernel3):

$$K(x, y) = \exp(-\gamma \times ||x - y||^2) \tag{3}$$

Features: have unlimited dimensions and can handle complex non-linear relationships. The width of the parameter gamma control function.

④ Sigmoid kernel function (kernel4):

$$K(x, y) = \tanh(\gamma \times x \times y + c) \tag{4}$$

Features: It comes from the neural network, with the S-shaped curve, which is suitable for certain specific problems.

ISSN 2616-5805 Vol. 5, Issue 3: 44-51, DOI: 10.25236/AJMS.2024.050307

2.2 GP regression based on multi-kernel learning and Bagging algorithm

The regression of the Gaussian process based on multi-kernel learning and Bagging algorithm combines the flexibility of multi-kernel learning and the integrated advantage of the Bagging algorithm.

In the first step, this article need to create a kernel function set: Let the kernel function set be $\{k|k_1, k_2, k_3 \ldots k_i, i = n\}$, each of which represents a basic kernel function, such as the RBF kernel, polynomial nucleus, etc. N is the number of nucleus functions.

Then the second step is to build a composite kernel function: The composite kernel function is constructed by combining the basic kernel function by combining the combination of method or multiplication. When combining combinations, the composite nucleus can be represented as:

$$K(x, x') = \sum_{n}^{1} \alpha_i k_i(x, x') \tag{5}$$

When the multiplication combination, the composite nucleus can be expressed as:

$$K(x, x') = \prod_{n}^{1} \alpha_i k_i(x, x') \tag{6}$$

In the third step, this article will train multiple GPR models: Based on the composite kernel function, this article train multiple GPR models independently on different training subsets. Each GPR model is predicted according to the assumptions of the Gauss process, that is, assuming the target function is extracted from a Gaussian process, and its nature is defined by the composite kernel function.

Finally, this article integrates predictions through the Bagging algorithm: After getting the prediction results of multiple GPR models, this article uses the Bagging algorithm to integrate them. Specifically, the final prediction output can be obtained by combining the prediction results of these models in average or weighted. The integration method of the Bagging algorithm improves the stability of the forecast and reduces the variance of the model.

3. Data analysis

3.1 Data source and experimental settings

The data source of this experiment mainly contains two parts. The first is the simulation data. The generating mechanism of the simulation data is shown below:

$$y = \sin(\frac{\pi}{2} \times (x_1 + x_7)) + \cos(\frac{2}{3}\pi \times x_2) + 2 \times (x_3 + x_4 + x_5 + x_6) + a \tag{7}$$

$$y = \sin(\frac{\pi}{2} \times (x_1 + x_7)) + \cos(\frac{2}{3}\pi \times x_2) + 2 \times (x_3 + x_5 + x_6) + a \tag{8}$$

$$y = \exp(x_1) + \exp(\frac{x_2}{2}) + 2 \times (x_3 + x_4 + x_5 + x_6 + x_7) + a \tag{9}$$

$$y = \exp(x_1) + \exp(\frac{x_2}{2}) + 2 \times (x_3 + x_5 + x_6 + x_7) + a \tag{10}$$

X comes from uniform distribution, a is a random term and set to a normal distribution. In order to explore the effects of different samples on the performance of the model, this article sets the sample volume n to 100 and 300, and perform 100 simulation experiments on each sample amount. And in the process of simulation, this article introduces different degrees of disturbance to simulate noise and uncertainty in the real world. In the end, this article will show the simulation results and empirical results through the chart to analyze the performance of the model more intuitively.

Real data is the source of Tushare platform file. Among them, 000001.SZ represents Ping An Bank Co., Ltd.,which belongs to the banking industry in the financial industry; 000002.SZ represents Vanke Enterprise Co., Ltd. is the leading enterprise in the real estate industry; 000004.SZ corresponds to Shenzhen Guoong Technology Co., Ltd. The company is mainly engaged in information transmission, software and information technology services; 000005.SZ is the stock code of Shenzhen Century Xingyuan Co., Ltd., which also belongs to the real estate industry. These data reflect the performance of these companies in the stock market and are an important basis for investors to study and decide.

In order to conduct rolling forecasts, this article sets data from February, March, and April to predict the rise and fall in May, and so on. In terms of model construction, this article selects a variety of kernel functions such as Gaussian kernel, linear kernel and polynomial nuclei to capture different characteristics in the data. In addition, this article uses the Bagging integration method, where the base model is selected as a multi-linear model to enhance the stability and predictive performance of the model.

The average method (MSE) is the average value of the predictive error, which measures the average difference between the predictive value and the actual value. The average deviation error (ABE) is the average value of the deviation between the prediction value and the actual value, which measures the average degree of deviation of the predictive value.The relative error efficiency (REE) is the measurement of the two model prediction errors, and is usually used to compare the model performance of different models or different parameter settings. The smaller the MSE, ABE and REE values are, the better the predictive performance of the model.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - y^\wedge{}_i)^2 \tag{11}$$

$$\mathrm{ABE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - y^\wedge{}_i| \tag{12}$$

$$\mathrm{REE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y^\wedge{}_i}{y_i}\right| \tag{13}$$

3.2 Comparison results

In this round of simulation experiments, this article adopted two different data generation methods to create analog dataset and each method considers all variable participation and some variable participation.In response to each data generation method and variable participation, this article further refines the setting of the sample volume, use 100 and 300 samples, and consider two parameter values of 0.01 and 0.09.

At the same time,this experiment compared the performance of different machine learning algorithms (including decision-making trees DTree, XGBoost, random forest RF, K close neighbor KNN, multi-layer perceptual MLP and Bagging_GP based on Gaussian process on the same dataset.By training and testing these algorithms, this article has calculated the mean (average performance) and std (standard deviation) of each algorithm to measure its stability and accuracy.

Based on the above factors, this article has conducted 16 different experimental combinations. By comparing these experimental results, this article found that the performance of different models shows a consistent law. That is, as the amount of samples increases, no matter what data generation method or variable participation, the error of the model shows a decrease. This shows that under the increasing data scale, the model can better capture the inherent laws and characteristics of data, thereby improving the accuracy of prediction. In addition, the experimental results also showed that when the parameter value increased from 0.01 to 0.09, the error of the model also increased. This shows that the changes in parameters have a significant impact on the performance of the model. In terms of average performance (mean), no matter what data generation method and sample amount or parameter, this article can see value, which means that it performs best in the average prediction accuracy. In terms of performance stability (std),the Bagging_GP algorithm also achieves minimum value, which means that its performance fluctuations are the smallest, that is, the prediction results are more stable and reliable. In contrast, other algorithms such as Dtree, KNN and MLP show a large fluctuation in performance stability. XGBoost and RF also perform well in terms of performance stability but it is slightly inferior to Bagging_GP, as shown in Table 1, Table 2, Table 3, and Table 4.

Table 1: The analog data result error analysis table based on the sample volume and parameters of the Model 1

Model 1	n=100		n=300	
	δ= 0.01	δ= 0.09	δ= 0.01	δ= 0.09
Bagging_GP	0.0041(0.0010)	0.0222(0.0091)	0.0008(0.0001)	0.0164(0.0027)
LASSO	1.6710(0.4885)	1.6175(0.4039)	1.6011(0.1333)	1.6799(0.2074)
DTree	1.1340(0.2240)	1.1406(0.3560)	0.7075(0.0982)	0.7619(0.1232)
SVR	0.2141(0.0968)	0.1943(0.0811)	0.0763(0.0258)	0.1009(0.0233)
RF	0.5181(0.1926)	0.4848(0.1102)	0.3153(0.0423)	0.3076(0.0551)
KNN	0.4354(0.0963)	0.4627(0.1335)	0.2835(0.0479)	0.2902(0.0464)
MLP	1.2831(0.3559)	1.3459(0.4280)	0.0237(0.0137)	0.0289(0.0073)

ISSN 2616-5805 Vol. 5, Issue 3: 44-51, DOI: 10.25236/AJMS.2024.050307

Table 2: The analog data result error analysis table based on the sample volume and parameters of the Model 2

Model 2	n=100		n=300	
	δ= 0.01	δ= 0.09	δ= 0.01	δ= 0.09
Bagging_GP	0.0149(0.0048)	0.0332(0.0122)	0.0185(0.0214)	0.0274(0.0155)
LASSO	1.4027(0.3398)	1.3685(0.1867)	1.4555(0.1417)	1.4412(0.2092)
DTree	0.8934(0.3749)	1.0590(0.2752)	0.7495(0.1016)	0.7240(0.0983)
SVR	0.1466(0.0562)	0.1360(0.0503)	0.0682(0.0165)	0.0806(0.0195)
RF	0.3883(0.0911)	0.4554(0.0652)	0.2861(0.0497)	0.2722(0.0548)
KNN	0.4061(0.1408)	0.3556(0.0760)	0.2527(0.0372)	0.2524(0.0396)
MLP	0.5627(0.5177)	0.4589(0.4924)	0.0295(0.0064)	0.0362(0.0112)

Table 3: Based on the analysis table of the simulation data result error of the Model 3 in different cases and parameters

Model 3	n=100		n=300	
	δ= 0.01	δ= 0.09	δ= 0.01	δ= 0.09
Bagging_GP	0.0007(0.0009)	0.0224(0.0080)	0.0003(0.0001)	0.0182(0.0027)
LASSO	1.9703(0.2359)	2.0769(0.3868)	2.0562(0.2156)	1.9470(0.2965)
DTree	1.5356(0.3093)	1.4808(0.4069)	1.1809(0.2357)	1.1984(0.2875)
SVR	0.2284(0.0622)	0.2849(0.1706)	0.1037(0.0332)	0.1176(0.0462)
RF	0.6455(0.1401)	0.7985(0.3058)	0.4242(0.0746)	0.4078(0.0439)
KNN	0.5566(0.1250)	0.6065(0.2329)	0.3246(0.0518)	0.3185(0.0833)
MLP	0.7022(0.1121)	0.6713(0.1491)	0.1567(0.1012)	0.1292(0.0733)

Table 4: The analog data result error analysis table based on the sample volume and parameters of the Model 4

Model 4	n=100		n=300	
	δ= 0.01	δ= 0.09	δ= 0.01	δ= 0.09
Bagging_GP	0.0008(0.0002)	0.0219(0.0056)	0.0003(0.0001)	0.0188(0.0053)
LASSO	1.7495(0.3833)	1.5397(0.3346)	1.5449(0.1766)	1.5623(0.1145)
DTree	1.1444(0.3132)	1.0546(0.1872)	0.6779(0.1166)	0.6931(0.1158)
SVR	0.2265(0.0568)	0.1951(0.0799)	0.0670(0.0203)	0.0886(0.0121)
RF	0.4991(0.1053)	0.4876(0.0799)	0.2494(0.0376)	0.2892(0.0506)
KNN	0.4548(0.1236)	0.4691(0.1154）	0.2429(0.0403)	0.2680(0.0343)
MLP	0.8412(0.1371)	0.7843(0.2650)	0.0439(0.1148)	0.0153(0.0033)

In summary, the Bagging_GP algorithm based on the Gaussian process has achieved optimal performance on both indicators of average performance and performance stability, showing its excellent prediction capabilities and stability on the data set. Although other algorithms also have good performance in some aspects, in general, the Bagging_GP algorithm is the best choice.

Below this article applies the six models to the four real stock data. Next, this article present the experimental results of two stocks. These two real data are derived from the Tushare platform for different error forms and folding line maps. As shown in the figures, Figures 1-2 and Tables 5-10 present the following content: From the analysis of different errors from Table 5-Table 10:By comparing the six models:decision-making tree DTree, XGBoost, random forest RF, K-nearest neighbor KNN, multi-layer perceptron MLP and Bagging_GP, all based on the Gaussian process, this article found that these models are accurate in predicting 000002.SZ and 000004.SZ. And this article found that the Bagging_GP algorithm performed well when predicting the stock,and its square errors, absolute errors and relative errors were better than other models.This result proves the effectiveness of the Bagging_GP algorithm on handling stock prediction issues, providing investors with more accurate and reliable decision support. Therefore, the Bagging_GP algorithm has broad application prospects in the field of stock forecasting. Based on the experimental results, a folding line diagram of absolute error, relative error and average square error were drawn, which intuitively shows the performance differences of different models when predicting specific stocks. It can be seen from the folding diagram that the Bagging_GP algorithm based on the Gaussian process maintains a low level on the three error indicators, and in most cases, it is better than the other five models.This further proves the stability and reliability of the Bagging_GP algorithm on the issue of stock prediction.

Table 5: The prediction square error of 000002.sz

	5	6	7	8	9	10	11	12
Bagging _GP	0.0465	0.0628	0.0824	0.0735	0.0175	0.0312	0.0498	0.0380
DTree	0.1540	0.1154	0.1022	0.1351	0.0758	0.4904	0.0732	0.5049
XGBoost	0.0806	0.0663	0.0736	0.1058	0.1036	0.1970	0.0774	0.4999
RF	0.1073	0.0750	0.0935	0.1022	0.1218	0.0795	0.0613	0.1335
KNN	0.0708	0.1026	0.0709	0.1011	0.0488	0.0275	0.0732	0.0654
MLP	0.1188	0.0913	0.0953	0.0894	0.1714	0.0382	0.1166	0.1871

Table 6: The prediction absolute error of 000002.sz

	5	6	7	8	9	10	11	12
Bagging_GP	0.1646	0.2051	0.1540	0.2119	0.1028	0.1390	0.1555	0.1553
DTree	0.3020	0.2858	0.2365	0.2706	0.2530	0.6729	0.2096	0.6888
XGBoost	0.2438	0.2575	0.2168	0.2572	0.2782	0.4124	0.2134	0.6858
RF	0.2991	0.2528	0.2437	0.2419	0.3094	0.2325	0.1654	0.3457
KNN	0.2285	0.2422	0.1544	0.2597	0.1597	0.1371	0.1933	0.2107
MLP	0.3206	0.2830	0.2274	0.2732	0.3581	0.1680	0.2299	0.3967

Table 7: The prediction relative error of 000002.sz

	5	6	7	8	9	10	11	12
Bagging_ GP	1.0109	1.0560	0.9436	0.9920	0.9923	1.0200	0.8843	0.9369
DTree	2.5359	2.7001	3.5925	2.4815	6.0637	10.8487	1.9318	6.2009
XGBoost	1.8631	2.2546	2.9747	2.0544	6.7654	6.1301	2.0247	6.1765
RF	2.4751	2.0470	3.7601	1.8179	7.7578	2.7449	1.2031	2.9957
KNN	1.7662	1.2734	2.0706	1.9475	2.5868	1.4733	1.7180	1.8181
MLP	2.7763	2.6195	3.0775	2.1610	8.9834	3.0374	2.2382	3.4693

Through experiments, the prediction of the real data of the two stocks is displayed, and this article finds that the Bagging-GP algorithm has achieved the minimum value on the three indicators of absolute error, relative error and average square error, which fully proves that it is in stock forecasting in stock forecasting excellent performance. In addition, this article have also drawn the folding line diagram corresponding to different errors to further show the stability and superiority of the Bagging-GP algorithm on the predictive error. The results of comprehensive experiments, this article can conclude that the Bagging-GP algorithm has a significant advantage in stock prediction issues. It not only has high prediction accuracy but also has good stability, providing investors with reliable decision support. Therefore, the Bagging_GP algorithm has broad application prospects in the field of stock forecasting.

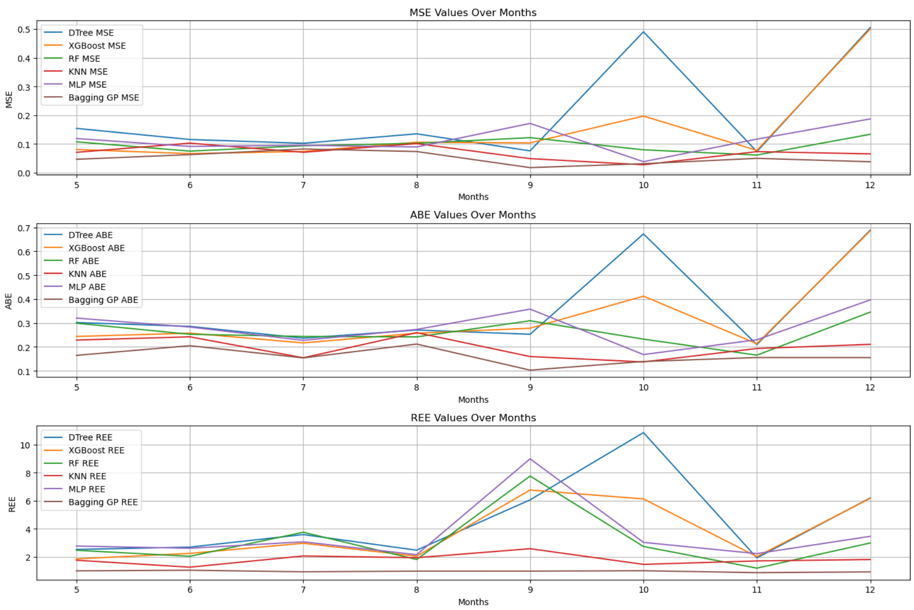

Figure 1: 000002.sz Real historical stock price prediction forecast for three error diagrams under six different models

ISSN 2616-5805 Vol. 5, Issue 3: 44-51, DOI: 10.25236/AJMS.2024.050307

Table 8: The prediction square error of 000004.sz

	5	6	7	8	9	10	11	12
Bagging _GP	0.1869	0.5377	1.9193	0.6432	0.2172	0.7433	0.2770	0.1178
DTree	0.3115	1.0617	3.0237	2.4047	0.8774	0.8600	0.8007	0.4866
XGBoost	0.4339	0.6738	2.8485	1.9521	0.4254	0.6519	0.3308	0.3762
RF	0.2224	0.5975	2.5460	1.0092	0.3854	0.6588	0.2751	0.1900
KNN	0.3059	0.6130	2.1239	1.3884	0.3840	0.6620	0.3640	0.2329
MLP	4.1807	0.9709	14.3097	1.0318	0.2637	0.8775	1.1745	0.3210

Table 9: The prediction absolute error of 000004.sz

	5	6	7	8	9	10	11	12
Bagging_GP	0.3681	0.6203	1.2575	0.6210	0.3081	0.6930	0.3857	0.2936
DTree	0.4939	0.7422	1.5901	1.3247	0.6493	0.7964	0.6537	0.5779
XGBoost	0.5693	0.6722	1.5187	1.0699	0.4311	0.6872	0.4955	0.5832
RF	0.4114	0.6439	1.4173	0.7727	0.3831	0.6342	0.4238	0.3713
KNN	0.4760	0.6747	1.2203	0.9796	0.4885	0.6149	0.4095	0.4090
MLP	1.4379	0.8722	3.3366	1.0263	0.3646	0.7934	0.9231	0.5001

Table 10: The prediction relative error of 000004.sz

	5	6	7	8	9	10	11	12
Bagging_ GP	0.9967	0.9536	1.0007	1.0461	1.5326	2.0908	1.0028	0.9990
DTree	2.2097	3.7808	1.3477	4.9011	4.7768	2.6950	2.7219	2.9267
XGBoost	2.8706	2.2946	1.2471	3.9380	3.0029	2.0732	1.8195	2.8391
RF	1.3978	1.5625	1.1108	2.2654	2.4736	1.7268	1.3384	1.5382
KNN	1.6344	2.4479	0.9089	2.3692	2.8564	1.0769	1.0636	1.7300
MLP	8.4405	4.6406	3.3305	3.4594	2.0561	2.7861	3.9424	2.5058

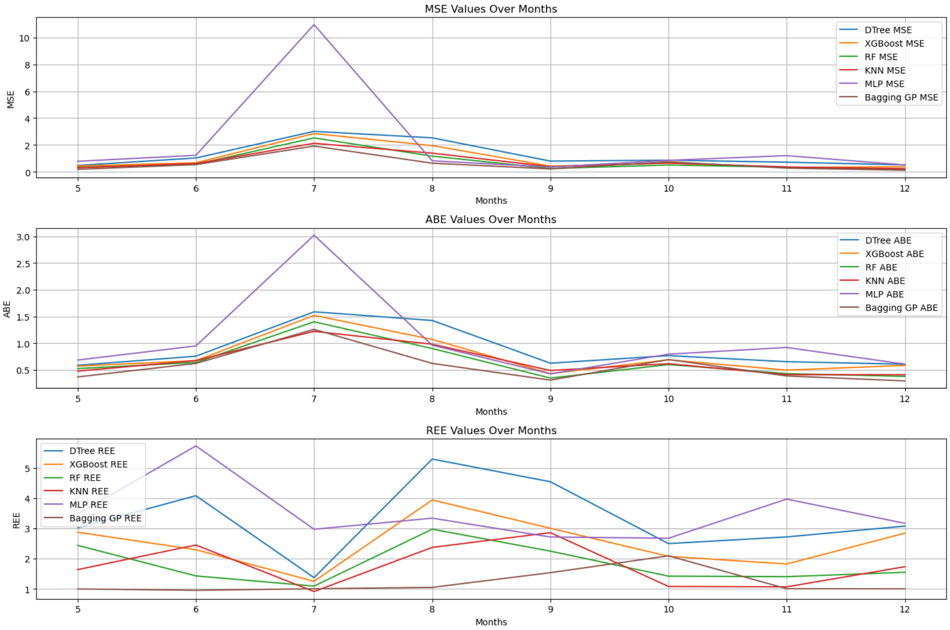

Figure 2: 000004.sz Real historical stock price predicting three error diagrams in six different models

4. Conclusion and prospects

This article proposes a self-adaptive Gaussian process regression model using a variety of kernel functions for a model, aiming to solve the problem of the Gaussian process regression and enhance the stability of the model. In the analog data analysis, the method of mentioning a smaller square error, relative error and absolute error showed a smaller method than other comparison methods, which verify

ISSN 2616-5805 Vol. 5, Issue 3: 44-51, DOI: 10.25236/AJMS.2024.050307

its superior performance. In practical applications, this article applies the method mentioned to the actual stock data of quantitative transactions and get satisfactory results. This model not only accurately predicts the trend of stock prices but also shows good stability when facing market fluctuations.

Regarding the future research direction, first of all, the selection and combination of kernel functions will be a complex problem. In the future, you can study more efficient kernel functions to select algorithms and combined strategies to further improve the performance of the model. Secondly, this article only considers the use of the Bagging algorithm to enhance the model. In the future, it can explore the possibility of combining other integrated learning methods (such as Boosting, Stacking, etc.) and multi-kernel learning. In addition, you can also consider applying the method mentioned in data analysis problems in other fields to expand its application scope. In short, the regression model of the adaptive Gaussian process mentioned in this article provides a new idea and method for solving the problem of kernel function selection.

References

[1] Wang S, Mo D, Wu Q, et al. Design and analysis of sustainable models for Qinling ecological protection and mining development[J]. Minerals Engineering, 2023, 204: 108446.
[2] He Y , Liu J N K , Hu Y ,et al.Fuzzy Integral-Based Kernel Regression Ensemble and Its Application[J].handbook of research on artificial intelligence techniques & algorithms, 2015.DOI:10.4018/978-1-4666-7258-1.ch012.
[3] Jones, O., Scheuerlein, A., Salguero-Gómez, R. Et Al. Diversity of Ageing across the Tree of Life. Nature 505, 169–173 (2014).
[4] Shu Shike, Li Lu. Multi -factor stock selection strategy based on Elastic Net punishment [J]. Statistics and decision -making, 2021 (16): 5.
[5] Yan Zhengxu, Qin Chao, Song Gang. Random forest model stock price prediction based on Pearson feature selection [J]. Computer engineering and application, 2021, 57 (15).
[6] Tian, L., FENG, L., yang, L. Et Al. Stock price prediction based on LSTM and LightGBM hybrid models. Super Computing Magazine 78, 11768–11793 (2022).
[7] Breiman Leo. Random Forests [J]. Machine Learning, 2001, 45 (1): 5-32.
[8] Wang Labin, Li Chao, Yang Qinmin, et al. A Wind Speed Vector Prediction Method Based on Gaussian Process and Combined Model: CN202211356272.6[P]. CN115907099A [2024-03-25].
[9] Yang Zhenli, Xia Kewen. The stock price prediction model regressioned based on the Gauss process [J]. Computer simulation, 2013, 30(1):5.
[10] Zhu Hongyu Gauss's process model prediction of stock prices [D]. Yunnan University, 2017.
[11] He Zhongzheng, Fang Li, Liu Wan, etc. The short-term runoff prediction research based on the index kernel function Gauss process [J]. China's rural water conservancy and hydropower, 2023 (8): 25-31.
[12] Zhu Shuxian, Zhang Renjie. Research on the selection of the kernel function of the vector machine [J]. Science and Technology and Engineering, 2008, 8(16): 5.
[13] Sun Na, Zhang Nan, Zhang Shuai, et al. A monthly runoff forecast model based on the Gaussian process of the Gaussian process and its application [J]. Hydropower and Energy Science, 2023.
[14] Yang Mengmeng, Wang Xinghui, Zhao Xing. Stock price forecast based on the average method of enhanced the model based on dynamic factor [J].Journal of Qilu University of Technology, 2022, 36(5):75-82.
[15] Qiu Shan, Gong Wenjie, Zhang Zhisheng. Based on Bagging-short-term load prediction model study based on the Bagging-combination kernel function related vector machine [J].Journal of Electrical Engineering, 2023, 18(2):142-148.
[16] Niu Hongli, Zhao Yazhi. Use the Bagging algorithm and GRU model to predict the stock price index [J]. Computer engineering and application, 2022, 58 (12): 132-138.