Research On Regression of High-Dimensional Redundant Data Based on Elastic Net and Gaussian Process

Changsheng Zheng [1, *], Xinyang Miao [2]

[1] School of Science, China University of Mining and Technology-Beijing, Beijing, China, 100083

[2] Hongyi Honor College, Wuhan university, Wuhan, China, 430072

* Corresponding Author Email: 18810399552@163.com

Abstract. In modern scientific research and engineering practice, managing high-dimensional and redundant datasets has become a prevalent challenge. Such datasets often contain a substantial amount of noise and irrelevant information, which significantly complicates traditional regression modeling approaches. Consequently, the development of a regression model that can effectively tackle high-dimensional redundant data, thereby extracting valuable information from complex datasets and making precise predictions, is of paramount importance. To tackle this challenge, this paper introduces an innovative regression model. The method integrates L1 and L2 regularization techniques to diminish redundant information within high-dimensional data, thereby enhancing the model's generalization capabilities. Simultaneously, the elastic net approach balances the model's complexity with noise sensitivity when handling high-dimensional data, preventing overfitting and underfitting, and thus improving prediction accuracy. The model parameters are determined using the maximum likelihood estimation algorithm. Simulation experiments and actual data analysis demonstrate that the proposed method surpasses benchmark methods in performance, showcasing its robust competitiveness.

Keywords: High-dimensional data, Elastic net, Gaussian process, Regression model, Feature selection.

1. Introduction

In modern scientific research and engineering practice, high-dimensional and redundant datasets are widely prevalent. Such data often contain a large amount of noise and irrelevant information, posing significant challenges to traditional regression modeling methods. In high-dimensional spaces, traditional regression models are prone to overfitting, leading to a decrease in the generalization ability of the model, and thus unable to effectively extract useful information from complex data and make accurate predictions. Therefore, how to develop regression models capable of effectively handling high-dimensional redundant data, and enhancing the predictive accuracy and generalization ability of regression models, has become an urgent research issue to be addressed. This paper aims to propose a novel regression model specifically designed to tackle the challenges posed by high-dimensional redundant data, and to provide reliable tools for practical applications.

In high-dimensional data analysis, the processing of redundant variables and complex nonlinear relationships are two core challenges. Redundant variables can lead to model overfitting and reduce the interpretability of the model, while the failure to effectively capture nonlinear relationships can affect prediction accuracy. In recent years, researchers have proposed various solutions to address these two issues. For instance, Tibshirani et al. (1996) used L1 regularization (Lasso) and L2 regularization (Ridge) for feature selection to reduce redundant information in high-dimensional data, thereby enhancing the generalization ability of the model [1-2]. Li Lijun et al. (2023) proposed a ReliefF feature selection algorithm based on redundancy analysis to solve the sample representativeness problem in traditional ReliefF algorithms, significantly improving classification accuracy in models such as LightGBM and SVM. Zhang Jing et al. (2023) proposed a feature selection method based on mutual information, which can effectively reduce feature dimensions, enhance the generalization ability of the model, and reduce computational complexity [3-4]. Zheng Jia and Li Shuyou (2022) used the maximum likelihood estimation method to estimate the parameters

of nonlinear regression models and perform data simulation, finding that adjusting the error term can effectively improve prediction accuracy [5-7]. Furthermore, regarding the feature selection problem of high-dimensional data, Zhang Tengfei et al. (2022) proposed an improved neighborhood rough set model, further enhancing the feature selection performance of high-dimensional data [8-10].

Despite the progress made in research on redundant feature selection and nonlinear regression modeling, existing methods still have some shortcomings. Elastic nets typically rely on manual feature engineering when dealing with nonlinear problems, while Gaussian processes have lower computational efficiency in large sample scenarios. Therefore, how to combine the feature selection ability of elastic nets with the nonlinear modeling ability of Gaussian processes has become a research hotspot. To this end, this paper proposes a high-dimensional redundant data regression model based on the fusion of elastic nets and Gaussian processes, aiming to effectively solve the problems of high-dimensional redundancy and nonlinear modeling. Through simulation experiments and real data analysis, this paper verifies the effectiveness and superiority of the proposed model in dealing with high-dimensional redundant data nonlinear problems, providing new insights and methods for high-dimensional data modeling.

2. Theory and Method

In the field of regression analysis, a general regression model is typically represented as:

$$Y = X^T\beta + \varepsilon_i \tag{1}$$

In which X represents the matrix of predictor variables, where the columns correspond to different predictors and the rows correspond to individual observations; β is the coefficient vector to be estimated; and ε is the random error term. However, in practical applications, X may have several issues. On one hand, X may contain variables that are unrelated to the response variable Y, which not only increases the complexity of the model but may also interfere with the model's ability to capture the true relationships; on the other hand, the variables in X may suffer from multicollinearity, meaning that there is a strong linear correlation between some variables, which can lead to unstable coefficient estimates and reduce the reliability and interpretability of the model. To address the issues of irrelevant variables and multicollinearity in X, methods such as LASSO (Least Absolute Shrinkage and Selection Operator) and Ridge regression are commonly employed. LASSO introduces an L1 penalty term into the objective function, which can result in some coefficient estimates being exactly zero, thereby achieving variable selection and automatically eliminating irrelevant variables. Ridge regression, on the other hand, incorporates an L2 penalty term, leading to "shrinkage" estimates of the coefficients, effectively alleviating the impact of multicollinearity and enhancing the stability of the coefficient estimates.

Simultaneously, the random error terms may also exhibit non-independence. When such a situation occurs, the assumptions of traditional regression models no longer hold, and Gaussian Process Regression (GPR) becomes an effective solution. The final regression model constructed using GPR is:

$$Y = X^T\beta + \varepsilon_i = X^T\beta + g(\mathrm{s}) + \varepsilon_i \tag{2}$$

In the aforementioned model, g(s) adheres to a Gaussian process with a mean of zero, and its variance is dictated by the kernel function K. The kernel function K serves to measure the similarity between various sample points. Commonly used kernel functions include the radial basis function and the polynomial kernel function, among others. Different kernel functions are appropriate for various data distributions and problem scenarios. Selecting the appropriate kernel function is essential for the model's performance. Thus, determining K and g(s) becomes the central task of model construction.

Let X denote the observed data, where X is the vector of predictor variables and Y represents the observed values of the response variable. The observational noise is independently and identically distributed, following a normal distribution with a mean of 0 and a variance of σ^2. Consequently, the response variable Y is also normally distributed:

$$Y \sim N(X^T\beta, K+\sigma^2 I) \tag{3}$$

Write $X^T\beta$ as μ, $K+\sigma^2 I$ as Σ .Meanwhile $X=(x_1,x_2,...,x_n)^T$, $Y=(y_1,y_2,...,y_n)^T$, $\mu=(\mu_1,\mu_2,...,\mu_n)^T$,Covariance matrix Σ:

$$\Sigma=\begin{pmatrix} K(x_1,x_1)+\sigma^2 & . & . & . & K(x_1,x_n) \\ . & & . & & . \\ . & & & . & . \\ . & & & . & . \\ K(x_n,x_1) & . & . & . & K(x_n,x_n)+\sigma^2 \end{pmatrix} \tag{4}$$

Next,seeking the likelihood function. The likelihood function represents the probability of the observed data occurring given the model parameters. For the aforementioned model, the likelihood function is:

$$L(\beta)=\prod_{i=1}^{n} p(Y_i=y_i)=\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(Y-\mu)^T\Sigma^{-1}(Y-\mu)\} \tag{5}$$

For the convenience of optimization, the negative log-likelihood function is usually calculated:

$$-L(\beta)=\frac{n}{2}\ln(2\pi)+\frac{1}{2}\ln|\Sigma|+\frac{1}{2}(Y-\mu)^T\Sigma^{-1}(Y-\mu) \tag{6}$$

To further address potential issues of irrelevant variables and multicollinearity in X, combining LASSO and Ridge regression, introducing two penalty functions, L1 and L2:

$$L1=\alpha_1\|\beta\|_1=\alpha_1\sum_{i=1}^{n}|\beta_i| \tag{7}$$

$$L2=\alpha_2\|\beta\|_2^2=\alpha_2\sum_{i=1}^{n}\beta_i^2 \tag{8}$$

Among them, α_1 and α_2 are hyperparameters used to control the intensity of the penalty. A larger α_1 will cause more coefficients to approach 0, enhancing the effect of variable selection; a larger α_2 will focus more on shrinking the coefficients, alleviating multicollinearity. These two hyperparameters typically need to be tuned using methods such as cross-validation to determine the optimal values for a given dataset. Define the new objective function $L_p(\beta)$ as the negative log-likelihood function plus the two penalty terms. Finally, gradient-based optimization algorithms (such as stochastic gradient descent, quasi-Newton methods, etc.) are used, with Python programming to solve:

$$\hat{\beta}=\arg\min L_p(\beta)=\arg\min(-L(\beta)+L1+L2) \tag{9}$$

Upon obtaining $\hat{\beta}$, since g(s) is typically difficult to obtain a specific expression and only its distribution can be determined, further calculations are needed to compute the distribution of the estimated response variable $\hat{y}$. According to the properties of Gaussian process regression:

$$\begin{pmatrix} y \\ \hat{y} \end{pmatrix} \sim N\left(\begin{pmatrix} X^T\hat{\beta} \\ x^{*T}\hat{\beta} \end{pmatrix}, \begin{pmatrix} K(X,X)+\sigma^2 I & K(X,x^*) \\ K(x^*,X) & K(x^*,x^*)+\sigma^2 I \end{pmatrix}\right) \tag{10}$$

As can be known from the properties of the conditional distribution of Gaussian process regression:

$$y_2 \mid y_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(y_1-\mu_1), \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) \tag{11}$$

Combining the aforementioned theory, by substituting actual data into the GPR model, the following can be obtained:

$$\begin{aligned} &\hat{y} \mid y \sim N(x^{*T}\hat{\beta} + K(x^*,X)[K(X,X)+\sigma^2 I]^{-1}(Y-X^T\hat{\beta}), K(x^*,x^*) \\ &+\sigma^2 I - K(x^*,X)[K(X,X)+\sigma^2 I]^{-1}K(X,x^*)) \end{aligned} \tag{12}$$

Finally, according to Bayesian estimation theory, the estimated value of the response variable y is:

$$\hat{y} = x^{*T}\hat{\beta} + K(x^*,X)[K(X,X)+\sigma^2 I]^{-1}(Y-X^T\hat{\beta}) \tag{13}$$

Through the comprehensive theoretical derivation and model construction process, by integrating the use of LASSO, Ridge regression, and Gaussian process regression, there are effectively addressed issues with predictor variables in general regression models, achieving accurate estimation and prediction of the response variables, and providing a powerful tool for practical data analysis and prediction tasks.

3. Data analysis

3.1. Simulation Experiment Analysis

In the simulation experiment, the independent variables are X and S, and the dependent variable is Y. Firstly, fixing the sample size of the independent variables X and S, their dimensionality, and their connection function with Y. Then, using the Monte Carlo method to randomly generate m sets of data with dimensionality n (and perform LASSO and Ridge regression processing on each set of data). Both X and S are random numbers drawn from the standard normal distribution, while Y is generated using the following formula (assuming the connection function for the nonlinear part is sin):

$$Y = X^T\beta + \sin(S) + \varepsilon \tag{14}$$

Where β represents a random vector of dimension n, drawn from a standard normal distribution, and ε is a random number from a normal distribution with a mean of 0 and a standard deviation of 0.1. Subsequently, divide the dataset into a training set (comprising 80%) and a test set (20%), and perform 10 experiments. Initially, train using the training set with the Elasticnet_GPR (Elastic Net Gaussian Process Regression model) presented in this paper, alongside five other distinct regression models (XGBRegressor, RandomForestRegressor, SVR, RBF, Elasticnet), to derive various regression coefficients and models. Following this, evaluate the Mean Squared Error (MSE) between Y_pred and Y_true using the test set. Ultimately, compute the Mean and Standard Deviation of the Mean Squared Error for each model. The Mean Squared Error's mean and standard deviation between Y_pred and Y_true serve as the evaluation metrics for this experiment: a reduced mean signifies a smaller model fitting error, while a reduced standard deviation indicates a superior model fitting performance. Furthermore, by adjusting experimental parameters (such as sample size, sample dimensionality, and the connection function with Y), how to configure model parameters to attain optimal fitting outcomes is obvious. The subsequent sections detail the experimental results when altering each of the three experimental parameters independently.

Initially, examining various sample sizes (100, 200, 300, 400), with the results depicted in Table 1 and Figure 1 below:

Table1. Mean Squared Error under Different Sample Sizes

Case1-1(100,10,sin)

Methods	Mean	Standard Deviation
Elasticnet-GPR	0.0134	0.0016
XGBoost	4.0323	1.6241
RF	3.1475	1.1289
SVR	3.4222	0.9533
GPR	1.5477	0.4391
Elasticnet	3.3659	1.0414

Case1-2(200,10,sin)

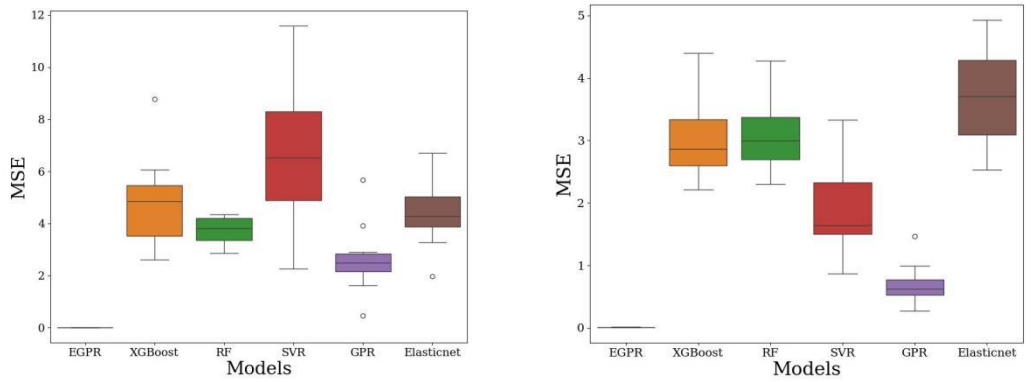
Methods	Method	Method
Elasticnet-GPR	0.0097	0.0023
XGBoost	3.0393	0.6358
RF	3.1126	0.5695
SVR	1.8644	0.7329
GPR	0.6917	0.3254
Elasticnet	3.7270	0.7607

Case1-3(300,10,sin)

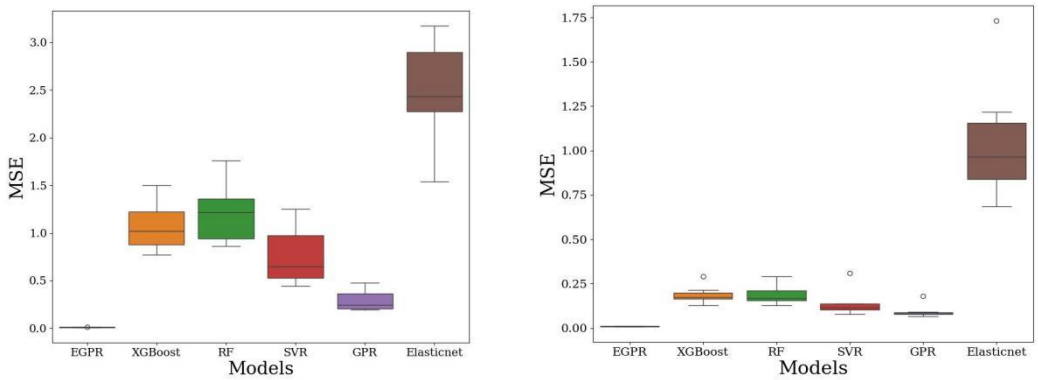
Methods	Mean	Standard Deviation
Elasticnet-GPR	0.0112	0.0019
XGBoost	1.0729	0.2283
RF	1.2031	0.2773
SVR	0.7454	0.2859
GPR	0.2879	0.1022
Elasticnet	2.4954	0.5009

Case1-4(400,10,sin)

Methods	Mean	Standard Deviation
Elasticnet-GPR	0.0103	0.0012
XGBoost	0.1838	0.0445
RF	0.1835	0.0463
SVR	0.1321	0.0623
GPR	0.0901	0.0310
Elasticnet	1.0397	0.2824



(a) Case1-1(100,10,sin) (b) Case1-2(200,10,sin)



(c) Case1-3(300,10,sin) (d) Case1-4(400,10,sin)

Figure 1. Mean Squared Error under Different Sample Sizes

Furthermore, considering the different dimensions of the predictive variable X and the number of irrelevant variables (10, 20, 30, 40), the results are shown in Table 2 and Figure 2:

Table 2. Mean Squared Error under Different Dimensions of X and Different Numbers of Irrelevant Variables

Case2-1(200,10,sin)

Methods	Mean	Standard Deviation
Elasticnet-GPR	0.0099	0.0012
XGBoost	1.4560	0.2726
RF	1.3138	0.3259
SVR	0.9124	0.2667
GPR	0.4628	0.1571
Elasticnet	2.9536	0.8301

Case2-2(200,20,sin)

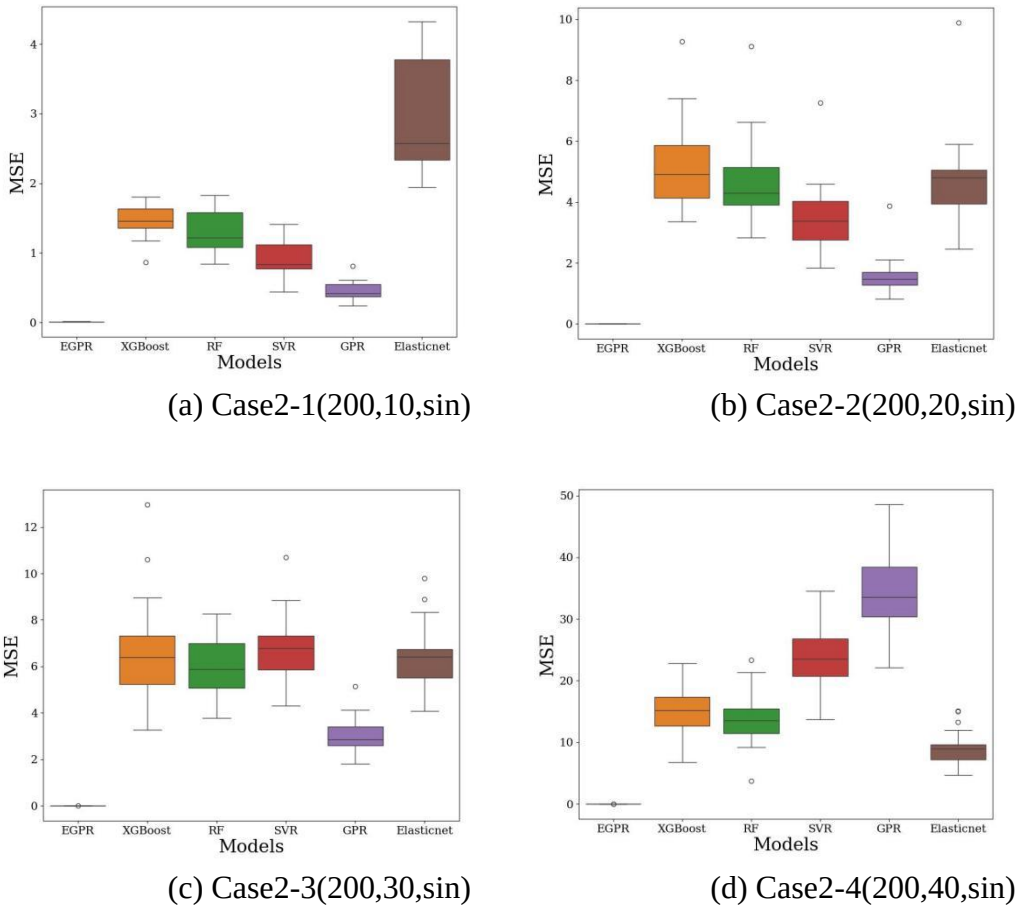
Methods	Mean	Standard Deviation
Elasticnet-GPR	0.0106	0.0020
XGBoost	5.1575	1.4124
RF	4.6078	1.3937
SVR	3.4311	1.2114
GPR	1.5637	0.6305
Elasticnet	4.6775	1.4982

Case2-3(200,30,sin)

Methods	Mean	Standard Deviation
Elasticnet-GPR	0.0113	0.0024
XGBoost	6.5841	1.9326
RF	5.9802	1.2190
SVR	6.7404	1.2996
GPR	2.9745	0.6761
Elasticnet	6.3436	1.2842

Case2-4(200,40,sin)

Methods	Mean	Standard Deviation
Elasticnet-GPR	0.0152	0.0044
XGBoost	15.0904	3.3214
RF	13.6558	3.6877
SVR	23.7025	4.9045
GPR	34.3131	6.3604
Elasticnet	8.9102	2.4233



(a) Case2-1(200,10,sin) (b) Case2-2(200,20,sin)

(c) Case2-3(200,30,sin) (d) Case2-4(200,40,sin)

Figure 2. Mean Squared Error under different dimensions of X and different numbers of irrelevant variables

Finally, considering different connection functions between S and Y, the results are shown in Table 3 and Figure 3:

Table 3: Mean Squared Error under Different Connection Functions with Y

Case3-1(200,10,sin)

Methods	Mean	Standard Deviation
Elasticnet-GPR	0.0139	0.0021
XGBoost	1.7885	0.3790
RF	1.6076	0.2724
SVR	1.5119	0.4204
GPR	0.6308	0.1647
Elasticnet	2.7991	0.4808

Case3-2(200,10,cos)

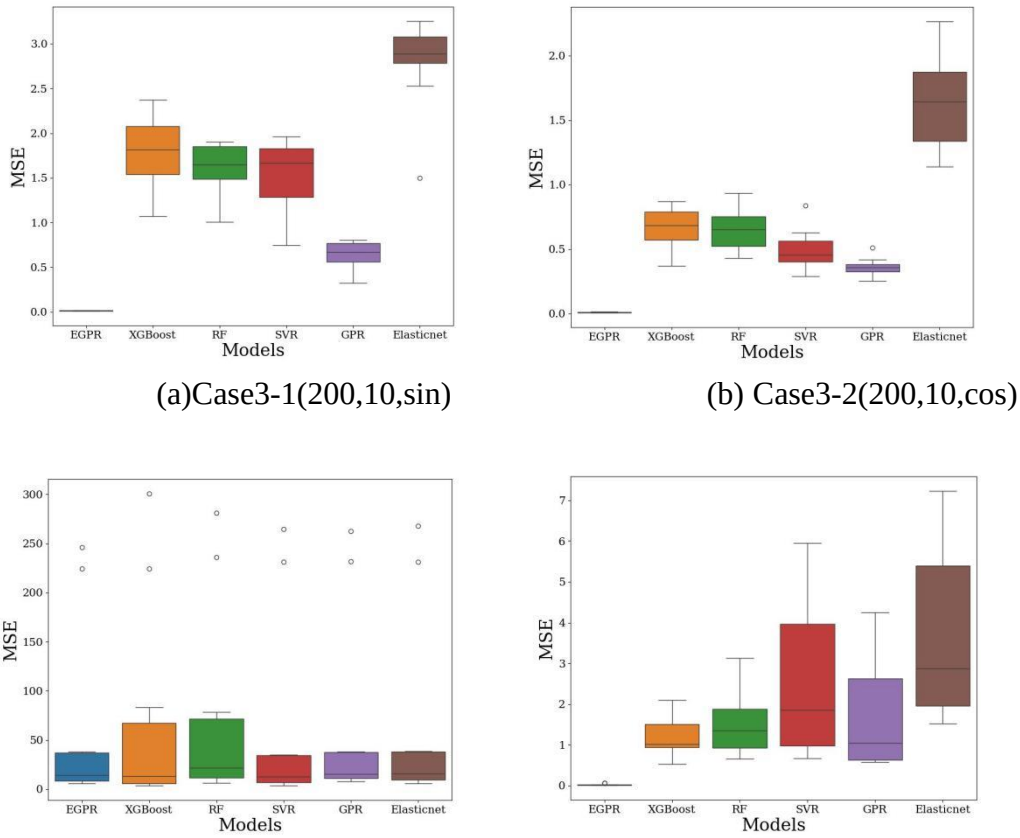
Methods	Mean	Standard Deviation
Elasticnet-GPR	0.0105	0.0020
XGBoost	0.6642	0.1560
RF	0.6480	0.1584
SVR	0.4952	0.1470
GPR	0.3597	0.0710
Elasticnet	1.6372	0.3378

Case3-3(200,10,tan)

Methods	Mean	Standard Deviation
Elasticnet-GPR	60.2057	88.3248
XGBoost	67.4636	101.4594
RF	72.8185	95.8986
SVR	61.3185	94.2091
GPR	64.0590	92.4211
Elasticnet	64.0356	93.6820

Case3-4(200,10,exp)

Methods	Mean	Standard Deviation
Elasticnet-GPR	0.0229	0.0181
XGBoost	1.2213	0.4894
RF	1.5013	0.7393
SVR	2.5073	1.8828
GPR	1.7261	1.4063
Elasticnet	3.6340	1.9771



(a)Case3-1(200,10,sin) (b) Case3-2(200,10,cos)

(c) Case3-3(200,10,tan) (d) Case3-4(200,10,exp)

Figure 3: Mean Squared Error under Different Connection Functions with Y

An in-depth analysis of the experimental results indicates that the EGPR model demonstrates exceptionally excellent characteristics as the sample size varies. Both the median and variance of the model are nearly zero, which strongly suggests that the model possesses outstanding fitting capabilities. In contrast, other regression models do not match the EGPR in terms of fitting effectiveness. Notably, as the sample size continues to grow, the fitting performance of each regression model generally exhibits a trend of gradual enhancement.

Further investigation reveals that when the dimension of X and the number of irrelevant variables change, the EGPR (Elasticnet_GPR) model maintains good stability, with its median and variance staying nearly zero, and the fitting performance remains very good. There are significant variations in the fitting performance of different models across various dimensions. For instance, the XGBoost (XGBRegressor) model's fitting performance initially declines and then gradually improves as the dimension increases; whereas the Elasticnet model steadily enhances with the increase in dimension.

Moreover, when the connection function with Y varies, the EGPR (Elasticnet_GPR) model also displays distinctive properties. Except when the connection function is tan, where the model's fitting effect is relatively poor, under other functions, it can achieve excellent fitting results. In comparison, other models, when the connection function is tan, may achieve a minimum median, but the variance does not necessarily reach a minimum.

Based on a comprehensive analysis of the experimental results from multiple perspectives, it can be concluded that when the sample size is moderate, the dimension of X and the number of irrelevant variables are also moderate, and the connection function is sin, cos, or exp, the EGPR (Elasticnet_GPR) model can exhibit very impressive fitting performance, distinguishing itself among numerous regression models.

3.2. Actual data analysis

In this section, applying the proposed EGPR (Elasticnet_GPR) optimal model to the spectral dataset of meat samples measured by a near-infrared spectrometer to analyze the performance of the EGPR (Elasticnet_GPR) optimal model in practical problems. The spectral dataset of meat samples can be downloaded from http://lib.stat.cmu.edu/datasets/tecator. This dataset contains n=240 samples, each with the content of moisture, fat, and protein in the meat, as well as the absorption spectra measured by the near-infrared spectrometer in the range of 850-1050 nanometers (nm). With an interval of every two bands, a total of 100 absorption spectral data points are recorded. Our goal is to predict the protein content of the meat samples based on their absorption spectral data, as well as the content of moisture and fat. Selecting samples from the dataset, where Y represents the protein content, s1 represents fat, s2 represents moisture, t represents the band ranging from 850-1050nm, and X(t) is the relatively easily measurable spectral curve. After assigning values, ten experiments are conducted, each dividing the dataset into a training set (80%) and a test set (20%). By substituting the training set data into the aforementioned simulation process, the regression coefficient estimates (β_est) of EGPR (Elasticnet_GPR) and five other regression models are obtained. The predicted values of Y (Y_pred) are then calculated using the remaining test set data (X_test, S_test). The mean and standard deviation of the mean squared error between the estimated values of Y and the true values of Y are then calculated, and the results are presented in Table 4 and Figure 4:

Table 4: Mean Squared Error of Meat Data

Methods	Mean	Standard Deviation
Elasticnet-GPR	0.7579	0.1664
XGBoost	1.5976	0.7182
RF	1.3950	0.5854
SVR	2.4442	0.8489
GPR	2.5912	0.7962
Elasticnet	2.3142	0.6989

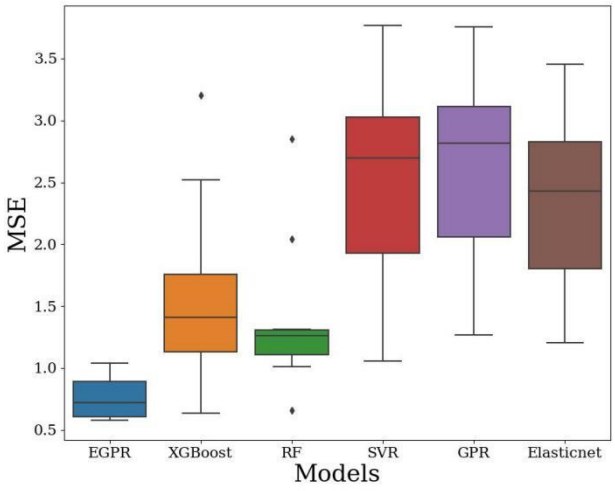


Figure 4: Mean Squared Error of Meat Data

From the chart information regarding the mean squared error of meat data, among the six regression models compared, the EGPR (Elasticnet_GPR) regression model stands out remarkably. Its median, mean, and standard deviation of the mean squared error (MSE) have all reached the lowest levels. Furthermore, the EGPR model's mean squared error data distribution is relatively concentrated, with a lower degree of dispersion. These data characteristics fully indicate that the EGPR (Elasticnet_GPR) regression model, when fitting meat data, can capture the patterns in the data more

precisely compared to the other five regression models. It demonstrates the best fitting effect and exhibits outstanding performance in regression analysis tasks.

4. Conclusion and Promotion

When addressing the modeling challenges of high-dimensional and redundant datasets, this paper innovatively proposes the EGPR (Elasticnet_GPR) regression model, which integrates Elastic Net and Gaussian Process Regression (GPR). This model fully harnesses the strengths of both Elastic Net and Gaussian Process Regression, effectively tackling the complex issues within the dataset. Elastic Net, by incorporating L1 and L2 regularization terms, can accurately pinpoint and discard irrelevant variables, while mitigating multicollinearity, significantly enhancing the stability and interpretability of the model. Gaussian Process Regression, with its distinctive kernel function mechanism, adeptly captures the nonlinear relationships within the data, effectively resolving the nonlinear challenges in regression models. The EGPR regression model boasts several remarkable advantages. On one hand, it can deeply explore high-dimensional data, efficiently reduce superfluous information, and thereby enhance the model's generalization ability, allowing it to maintain robust adaptability across various data distributions. On the other hand, the model balances complexity and noise sensitivity through meticulous tuning, cleverly sidestepping overfitting and underfitting, and markedly improving the model's prediction accuracy. Experimental results strongly validate the EGPR regression model's exceptional performance. In most experimental scenarios, it demonstrates superior fitting effects compared to traditional regression models. This fully underscores that the EGPR regression model is a viable option and should be considered for application in solving a variety of real-life problems. Additionally, the research findings of this paper offer ample scope for further exploration and can be delved into from the following two perspectives:

First, consider employing partial functional linear models as an alternative to Gaussian Process Regression to address the nonlinear challenges in regression models. Partial functional linear models merge the interpretability of linear models with the adaptability of nonlinear models, promising to introduce new insights and methodologies for managing complex data relationships.

Second, leveraging ensemble learning techniques, integrate six distinct models (including EGPR and the aforementioned comparative models) in an arithmetic mean fashion to create a new ensemble model, achieving comprehensive optimization of the regression model. By synthesizing the predictions from multiple models, the bias and variance of individual models can be substantially diminished, markedly enhancing the overall performance of the model. Owing to the synergistic effect of multiple models, the ensemble model exhibits enhanced resilience against the risk of overfitting inherent in individual models, possessing superior generalization capabilities, demonstrating consistent and outstanding performance across various datasets, and paving new research avenues and application possibilities within the realm of regression analysis.

References

[1] Robert T .Regression Shrinkage and Selection Via the Lasso[J].Journal of the Royal Statistical Society: Series B (Methodological),2018,58(1):267-288.

[2] Friedman J ,Hastie T ,Tibshirani R .Regularization Paths for Generalized Linear Models via Coordinate Descent[J].Journal of Statistical Software,2010,33(1):21-22.

[3] Li Lijun, Zhang Haiqing, Li Daiwei, et al. An Improved ReliefF Feature Selection Algorithm Based on Redundancy Analysis [J]. Software Engineering, 2023, 26(11):48-51.

[4] Zhang Jing, Cao Feng, Dong Yuying, et al. Feature Selection Algorithm Based on Mutual Information and Genetic Algorithm [J]. Journal of Shanxi University (Natural Science Edition), 2024, 47(01):1-8.

[5] Zheng Jia, Li Shuyou. Research on the Maximum Likelihood Estimation Method for Parameters of Nonlinear Regression Models Based on Uncertainty Theory[J]. Journal of Liaoning University of Technology (Natural Science Edition), 2023, 43(01): 24-28.

[6] Hu Guiping. A Case Study on Linearizing Nonlinear Regression Analysis Models[J]. Mathematics, Physics, and Chemistry Learning, 2024(20):9-14.

[7] Wei Dong. Non-Nuclear Nonlinear Regression Machine[D]. Xinjiang: Xinjiang University, 2023.

[8] Zhang Tengfei, Zhang Yudi, Ma Fumin. High-Dimensional Hybrid Data Feature Selection Algorithm Based on Improved Neighborhood Space[J]. Control and Decision, 2024, 39(3):929-938.

[9] Wang Zhenfei, Yuan Peiyao, Cao Zhongya, et al. Feature Selection Algorithm for High-Dimensional Imbalanced Data[J]. Journal of Mini-Micro Systems, 2024, 45(8):1839-1846.

[10] Yang Xuan. Feature Selection for High-Dimensional Time Series Data Based on Two Types of Metric Functions[D]. Shaanxi: Chang'an University, 2023.