A computing method of predictive value based on fitting function in linear model

Hao Zhong¹*, Huibing Zhang², Fei Jia²

¹School of Computer Science, South China Normal University, Guangzhou 510631, China
²Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China

Abstract

Linear models are common prediction models in collaborative computing, which mainly generates fitting function to express the relationship between feature vectors and predictive value. In the process of computing the predictive value according to the fitting function and feature vector, this paper mainly conducted the following researches. Firstly, this paper defines a change interval of predictive value according to training set. Secondly, in this paper, the change interval of predictive value corresponding to feature vector in test set is computed. Finally, according to distribution of training set in the changing interval, the predictive values corresponding to feature vectors in test set are computed. Standard data sets are used in experiment, and MAE(Mean Absolute Error) and RMSE(Root Mean Square Error) are used to evaluate the prediction results. The experimental results show that the method proposed in this paper can improve the prediction error to a certain extent.

Received on 07 June 2020; accepted on 23 September 2020; published on 02 October 2020

Keywords: linear model, linear fitting, fitting function, predictive value

1. Introduction

Prediction models in collaborative computing include linear models and non-linear models. Non-linear models include RF(Random Forest), NN(Neural Network), GBDT(Gradient Boosting Decision Tree), XGBoost(eXtreme Gradient Boosting) and so on. Linear models include linear regression, Ridge Regression, Lasso Regression and so on. The main advantage of linear models is that they are easy to model. The most important technology of linear model is linear fitting. In the process of prediction based on linear model, the fitting function generated according to the training set is called linear fitting. The process of linear fitting is as follows: a loss function is defined by training data, and some parameters of the fitting function are obtained when the loss function takes the minimum value. The parameters characterize the quantitative relationship between the feature vector and the predictive value, so the predictive value corresponding to feature vector of test data can be computed according to these parameters. According to the error between predictive value and true value of test data, the quality of predictive result is evaluated.


The above research work introduces various applications of linear models. Generally, linear models can be used in fields or directions that need data analysis, such
as Underwater Wireless Sensors Network[7], Underwater Routing[8] and Underwater Resurrection Routing Synergy[9]. The purpose of this paper is to optimize linear models and reduce predictive errors. Based on the parameters of fitting function, this paper proposes a method for computing a new predictive value, which can improve the accuracy of predictive result. The main contributions are as follows:

- Based on fitting function in linear model, a change interval of predictive value is defined. The change interval represents the potential change range of predictive value, which can better express the relationship between feature vector and predictive value.
- Based on the change interval of predictive value, a new computing method of predictive value is proposed. Experiments show that, new predictive value is closer to the real value than original prediction value.

The remainder of the paper is organized as follows. Section 2 introduces related work. Section 3 describes computing method briefly. Section 4 describes the process of computing method in detail. Section 5 describes the experimental results on the standard dataset. Section 6 is the summary of the paper.

2. Related Works

Linear Regression is the basis of all linear models. For the optimization of Linear Regression, the main direction is to optimize the solution process of loss function and improve convergence speed. The methods used mainly include Least Square Method, Batch Gradient Descent[10], Random Gradient Descent[11] or Mini-Batch Gradient Descent[12]. In addition, Zhang et al. [13] proposed a method of feature weighting, by giving important features a greater weight and minor features a relatively smaller weight, the fitted curve is closer to data distribution. Wang et al. [14] considered that high-dimensional features are more suitable for regression models, and proposed combining GBDT with regression models. He used GBDT to improve the dimensionality of input features and prediction accuracy.

By transforming loss function of Linear Regression, various linear models are generated and applied in different scenarios. For example, for the over-fitting problem in Linear Regression, Ridge Regression [15] added regularization term of L2 norm to loss function, which reduces regression coefficient of sparse features and improves stability of Linear Regression. Aiming at the problem these are large number of sparse features in high-dimensional data, Lasso Regression [16] added regularization term of L1 norm to loss function. It can delete some features which has small regression coefficient, reduce the dimension of features, and improve generalization ability of Linear Regression. Elastic-Net [17] combines the advantages of Ridge Regression and Lasso Regression, and adds a regularization term to loss function. By setting parameters to balance the proportion of the L1 regularization term and the L2 regularization term, the Elastic-Net results tend to Ridge Regression or Lasso Regression as the parameters change. Elastic-Net improves the stability of Linear Regression while improving the generalization ability of Linear Regression to some extent. Orthogonal Matching Pursuit [18] adds a restriction term to loss function, which can limit the maximum number of non-zero elements in the regression coefficient. Under a specified number of non-zero elements, it tends to get optimal regression coefficient to improve prediction accuracy of Linear Regression. Bayesian Ridge [19] assumes that the distribution of regression coefficients is spherical Gauss distribution, and estimates the regression coefficients by maximizing the marginal likelihood function. ARD Regression [20] assumes that the distribution of regression coefficients is elliptical Gauss distribution parallel to coordinate axis, and estimates the regression coefficients by maximizing the marginal likelihood function. When using loss function to calculate errors in samples, Huber Regressor [21] sets a threshold. When the error value of sample exceeds the threshold, sample is divided into abnormal data, and the abnormal data is given a small weight. To some extent, the stability of Linear Regression has been improved. In the process of finding the minimum value of loss function, Least Angle Regression [22] solves angle of current loss residual, and takes feature vector with smallest angle as a direction to find the minimum value. It can obtain ideal results after a limited number of iterations, increase speed at which loss function converges. Lasso Lars [23] combines advantages of Least Angle Regression and Lasso Regression. The regularization term of the L1 norm is added to loss function, and angle for computing current loss residual is found in the training set, and the feature vector with the smallest angle is taken as the direction for finding the minimum value. Improve the generalization ability of Linear Regression, and also increase the speed of the loss function convergence. Passive Aggressive Regressor [24] is a linear model proposed for big data, which incorporates the idea of perceptron into minimization process of loss function. However, the regularization parameter is added to perceptron, which limits parameter size of the perceptron, and avoids the over-fitting problem caused by rough classification of perceptron.

In view of loss function of linear regression, a series of modifications and improvements have been made in the current research work. The stability of the model, the
generalization ability of the model, and the convergence speed of loss function have been improved. Based on the fitting functions generated by various regressions, this paper redefines a computing method of prediction value without changing the algorithm complexity of linear model, in order to reduce predictive error.

3. Introduction of computing method

The definitions of symbols in this paper are shown in Table 1.

As shown in Algorithm 1, the process of the computing method is divided into the following four parts. The first step is training process of linear model. According to feature vectors $TRx = \{TRx_1, TRx_2, ..., TRx_n\}$ and corresponding true values $TRy = \{TRy_1, TRy_2, ..., TRy_n\}$, fitting function $Y = W^TX + \omega$ is generated. According to fitting function, the predictive value of feature vector $TRx_n$ is calculated and a set $(TRy_1, TRy_2, ..., TRy_n)$ is obtained. According to parameter $\alpha$, a potential change interval $[TRy_n^*\alpha, TRy_n^*(2-\alpha)]$ is defined, where original predictive value $TRy_n$ is the center of the interval.

In the second step, for training data, true values of the feature vectors are known. When $TRy_n$ is greater than $TRy_n^*(2-\alpha)$, $TRy_n^*(2-\alpha)$ is new predictive value corresponding to feature vector $TRx_n$. When $TRy_n$ is in interval $[TRy_n^*\alpha, TRy_n^*(2-\alpha)]$, $TRy_n$ is new predictive value corresponding to feature vector $TRx_n$, when $TRy_n$ is less than $TRy_n^*\alpha$, $TRy_n^*\alpha$ is new predictive value corresponding to feature vector $TRx_n$. When parameter $\alpha$ changes in interval $[0,1]$, the MAE and RMSE values of training data change accordingly. When MAE and RMSE take the minimum value, parameter $\alpha$ is determined.

In the third step, predictive values $\{TEy.P_1, TEy.P_2, ..., TEy.P_m\}$ of test set $TEx = \{TEx_1, TEx_2, ..., TEx_n\}$ is computed according to fitting function, and change interval $[TEy.D_m, TEy.U_m]$ of each predictive value $TEy.P_m$ is obtained from parameter $\alpha$. For each change interval $[TEy.D_m, TEy.U_m]$, according to original predictive value $TRy_n$ and true value $TRy_n$, of each training data, the distribution $TRx^{'} = \{TRx_1^{'}, TRx_2^{'}, ..., TRx_1^{'}\}$ of training data in the interval is generated.

In the fourth step, for test data $TEx_m$, the following known conditions include original predictive value $TEy.P_m$, change interval $[TEy.D_m, TEy.U_m]$ of predictive value $TEy.P_m$, and training data distribution $TRx^{'} = \{TRx_1^{'}, TRx_2^{'}, ..., TRx_1^{'}\}$ in interval $[TEy.D_m, TEy.U_m]$. According to these three known conditions, the final predictive value $TEy.F_m$ of $TEx_m$ is computed.

4. Computing method and Complexity Analysis

Algorithm 1 The computing process of new prediction value in linear model

Input: The feature vector of training set $TRx = \{TRx_1, TRx_2, ..., TRx_n\}$, the feature vector of test set $TEx = \{TEx_1, TEx_2, ..., TEx_m\}$, and the true value of training set $TRy = \{TRy_1, TRy_2, ..., TRy_n\}$

Output: The new predictive value $TEy.F_m$ of $TEx_m$

1: Getlossfunction($TRx, TRy$); // By the training set
2: Get $W^T, \omega$; // When the loss function takes the minimum value
3: Define the potential change interval $[TRy_n^*\alpha, TRy_n^*(2-\alpha)]$; // By $W^T$ and $\omega$
4: $\alpha \leftarrow \text{Getparameter}(TRy_n^*\alpha, TRy_n^*(2-\alpha), TRy_n)$
5: $TRx^{'} \leftarrow \text{Getlabels}(TEy.P_m^*\alpha, TEy.P_m^*(2-\alpha), TRy, TRy_P, TRy)$
6: $TEy.F_m \leftarrow \text{Getparevalue}(TEy.P_m^*\alpha, TEy.P_m^*(2-\alpha), TRx^{'})$

4.1. Change interval of predictive value

In training process of linear model, according to feature vectors of training sets $TRx = \{TRx_1, TRx_2, ..., TRx_n\}$, parameters $W^T$ and $\omega$ are set, and loss function is defined. By solving the minimum of loss function, parameters $W^T$ and $\omega$ are obtained, the fitting function $Y = W^TX + \omega$ is generated. According to fitting function, the predictive values $\{TRy_1, TRy_2, ..., TRy_n\}$ of all training data are computed. Taking each predictive value $TRy.n$ as center of interval $[TRy_n^*\alpha, TRy_n^*(2-\alpha)]$. At this time, the paper has the following definitions:

- If true value $TRy.n$ of $TRx_n$ is greater than $TRy_n^*(2-\alpha)$, set $TRy_n^*(2-\alpha)$ as predictive value of $TRx_n$.
- If true value of $TRy_n$ lies in interval $[TRy_n^*\alpha, TRy_n^*(2-\alpha)]$, set $TRy_n$ as predictive value of $TRx_n$.
- If true value $TRy_n$ of $TRx_n$ is less than $TRy_n^*\alpha$, set $TRy_n^*\alpha$ as predictive value of $TRx_n$.

Suppose the value of $TRy_n^*\alpha$ is represented by $a$, and value of $TRy_n^*(2-\alpha)$ is represented by $b$. The specific method is shown in formula (1).

$$
TRy.F_n = \begin{cases}
  b, & TRy_n \geq b \\
  \frac{a+b}{2}, & TRy_n \in (a, b) \\
  a, & TRy_n \leq a.
\end{cases}
$$

As shown in formula (1), since true value $TRy = \{TRy_1, TRy_2, ..., TRy_n\}$ of training set is known, according to computing of MAE and RMSE, the objective functions are defined as shown in formula (2) and formula (3).
Table 1. Symbols used in the paper

| Feature vectors of training set | $TR_x = \{TR_{x1}, TR_{x2}, ..., TR_{xn}\}$ |
| Feature vectors of test set | $TE_x = \{TE_{x1}, TE_{x2}, ..., TE_{xm}\}$ |
| True values of training set | $TR_y = \{TR_{y1}, TR_{y2}, ..., TR_{yn}\}$ |
| True values of test set | $TE_y = \{TE_{y1}, TE_{y2}, ..., TE_{ym}\}$ |
| Fitting function | $Y = W^T X + \omega$ |
| Original predictive value of $TR_x$ | $TR_y.P_n$ |
| Change interval value of $TR_y.P_n$ | $[TR_y.D_n, TR_y.U_n]$ |
| New predictive value of $TR_x$ | $TR_y.F_m$ |
| Original predictive value of $TE_x$ | $TE_y.P_m$ |
| Change interval value of $TE_y.P_m$ | $[TE_y.D_m, TE_y.U_m]$ |
| Distribution in $[TE_y.D_m, TE_y.U_m]$ | $TR_x' = \{TR_{x1}', TR_{x2}', ..., TR_{x'n}'\}$ |
| Location label of $TR_x'$ | $L = \{l_1, l_2,..., l_d\}$ |
| New predictive value of $TE_x$ | $TE_y.F_m$ |

$$\text{MAE}_. \text{TR} = \frac{1}{n} \sum_{i=1}^{n} |TR_y.i - TR_y.F_m| \quad (2)$$

$$\text{RMSE}_. \text{TR} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (TR_y.i - TR_y.F_m)^2} \quad (3)$$

The range of parameter $a$ is $(0,1)$. When both $\text{MAE}_. \text{TR}$ and $\text{RMSE}_. \text{TR}$ are minimum, parameter $a$ is obtained, that is, each test data $TE_y.P_m$ corresponds to a change interval $[TE_y.D_m, TE_y.U_m]$. The significance of defining the change interval is that, if it can correctly estimate the relationship between true value $TE_y.m$ and change interval $[TE_y.D_m, TE_y.U_m]$ (three relationships: $TE_y.m \geq TE_y.U_m$, $TE_y.m \leq TE_y.D_m$, $TE_y.m \in (TE_y.D_m, TE_y.U_m)$), it can get the minimum prediction error.

4.2. Computing method for new predictive values

After input test data $TE_x.m$, original predictive value $TE_y.m$ is computed according to fitting function $Y = W^T X + \omega$, and change interval $[TE_y.D_m, TE_y.U_m]$ is obtained according to parameter $a$. In training set, data $TR_x$ satisfying $TE_y.D_m \leq TR_y.P_n \leq TE_y.U_m$ is added to set $TR_x' = \{TR_{x1}', TR_{x2}', ..., TR_{x'n}'\}$, and corresponding location label $L_d$ of $TR_x'$ is shown in formula (4):

$$L_d = \begin{cases} 
2, & TR_y.n \geq TE_y.U_m, \\
1, & TR_y.n \in (TE_y.D_m, TE_y.U_m), \\
0, & TR_y.n \leq TE_y.D_m.
\end{cases} \quad (4)$$

According to formula (4), the number of 0, 1 and 2 in location label set $L = \{l_1, l_2,..., l_d\}$ is expressed by $S0$, $S1$ and $S2$ respectively. The larger value of $S2$, the greater probability that true value $TE_y.m$ of $TE_x.m$ satisfies $TE_y.m \geq TE_y.U_m$. The larger value of $S1$, the greater probability that true value $TE_y.m$ of $TE_x.m$ satisfies $TE_y.m \in (TE_y.D_m, TE_y.U_m)$. The larger value of $S0$, the greater probability that true value $TE_y.m$ of $TE_x.m$ satisfies $TE_y.m \leq TE_y.D_m$. In practice, the relationship between true value $TE_y.m$ and change interval $[TE_y.D_m, TE_y.U_m]$ cannot be accurately judged. With three values of $S0$, $S1$ and $S2$ as probabilistic values, the final predictive values $TE_y.F_m$ corresponding to $TE_x.m$ is defined as shown in formula (5):

$$TE_y.F_m = \frac{S2 \cdot TE_y.U_m + S1 \cdot TE_y.P_m + S0 \cdot TE_y.D_m}{S2 + S1 + S0} \quad (5)$$

4.3. Complexity analysis

The most important part of the linear model is to find the minimum of the loss function. Several main methods are Least Squares Method, Random Gradient Descent, Batch Gradient Descent and Mini-Batch Gradient Descent. The complexity of the linear model is the same as method for finding the minimum of the loss function. Assuming that data quantity is $N_f$ and dimension of feature vector is $N_f$, where $N_f >> N_f$. The complexity of least square method is $O(N_f \cdot N_f^2) = O(N_f^3)$. The complexity of Random Gradient Descent is $O(N_f \cdot N_f) = O(N_f^2)$. The complexity of Batch Gradient Descent and Mini-Batch Gradient Descent are $O(N_f^2)$.

According to the above analysis, the minimum complexity of linear model is $O(N_f)$. Based on fitting function of linear model, a method for calculating the new predictive value is proposed in this paper. The complexity of computing the change interval is $O(N_f)$. The complexity of obtaining data distribution is $O(N_f)$. The complexity of computing predictive value is $O(1)$. Generally speaking, the total complexity of the proposed method is $O(N_f) + O(N_f) + O(N_f) + O(1) = O(N_f)$, which improves the prediction effect of the model without changing complexity of the linear model.

5. Data sets and Experiments

5.1. Data Sets

In the sklearn database[25], three standard data sets are selected to test the effectiveness of the proposed
method. The data set is suitable for regression tasks. The selected data set and invocation method are shown in Table 2.

**Table 2. Datasets**

<table>
<thead>
<tr>
<th>Invoke Method</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>datasets.load-boston()</td>
<td>boston</td>
</tr>
<tr>
<td>datasets.fetch-california-housing()</td>
<td>california</td>
</tr>
<tr>
<td>datasets.load-diabetes()</td>
<td>diabetes</td>
</tr>
</tbody>
</table>

5.2. Experimental result of data sets

In the first experiment of the paper, Linear Regression (represented by LR) was used to predict and analyze the three data sets. At the same time, according to the method proposed in this paper, a new predictive value of Linear Regression (expressed by LR+) is computed. The predictive results were evaluated by using MAE and RMSE. The experimental results are shown in Table 3:

**Table 3. Prediction results of three data sets**

<table>
<thead>
<tr>
<th>data sets</th>
<th>a</th>
<th>LR (MAE)</th>
<th>LR+ (MAE)</th>
<th>LR (RMSE)</th>
<th>LR+ (RMSE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>california</td>
<td>0.74</td>
<td>0.54</td>
<td>0.52</td>
<td>0.73</td>
<td>0.72</td>
</tr>
<tr>
<td>boston</td>
<td>0.865</td>
<td>3.84</td>
<td>3.66</td>
<td>5.78</td>
<td>5.72</td>
</tr>
<tr>
<td>diabetes</td>
<td>0.724</td>
<td>46.17</td>
<td>45.18</td>
<td>58.52</td>
<td>58.07</td>
</tr>
</tbody>
</table>

As shown in Table 3, this paper presents a new method to compute the predictive value, which can reduce the predictive error to a certain extent without changing the complexity. This is because when test data are input, corresponding original predictive value is computed on fitting function. In practice, the true value of test data is located above or below original predictive value. If the position of true value can be correctly judged, prediction error can be reduced. In this paper, we take the original predictive value as a center, and define change interval of predictive value. According to data distribution in this interval, the position relationship between true value of test data and original predictive value is judged. New predictive values are computed according to the change interval and position relationship. The experimental results show that the new predictive value is closer to the true value.

5.3. Experimental result of linear models

In the second experiment of this paper, the data set boston was analyzed by linear models such as Ridge Regression and Lasso Regression. At the same time, according to the method proposed in the paper, computer new predictive values for linear models such as Ridge Regression+ and Lasso Regression+. MAE and RMSE are used to evaluate the predictive effect. The experimental results are shown in Table 4:

**Table 4. Prediction results of three data sets**

<table>
<thead>
<tr>
<th>line models</th>
<th>a</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ridge Regression</td>
<td>0.864</td>
<td>3.84</td>
<td>5.79</td>
</tr>
<tr>
<td>Ridge Regression+</td>
<td>0.864</td>
<td>3.65</td>
<td>5.72</td>
</tr>
<tr>
<td>Lasso Regression</td>
<td>0.836</td>
<td>4.20</td>
<td>6.32</td>
</tr>
<tr>
<td>Lasso Regression+</td>
<td>0.836</td>
<td>3.99</td>
<td>6.25</td>
</tr>
<tr>
<td>ElasticNet</td>
<td>0.833</td>
<td>4.34</td>
<td>6.53</td>
</tr>
<tr>
<td>ElasticNet+</td>
<td>0.833</td>
<td>4.15</td>
<td>6.51</td>
</tr>
<tr>
<td>OrthogonalMatchingPursuit</td>
<td>0.816</td>
<td>4.40</td>
<td>6.46</td>
</tr>
<tr>
<td>OrthogonalMatchingPursuit+</td>
<td>0.816</td>
<td>4.17</td>
<td>6.31</td>
</tr>
<tr>
<td>BayesianRidge</td>
<td>0.874</td>
<td>3.93</td>
<td>5.97</td>
</tr>
<tr>
<td>BayesianRidge+</td>
<td>0.874</td>
<td>3.73</td>
<td>5.87</td>
</tr>
<tr>
<td>ARDRegression</td>
<td>0.847</td>
<td>4.05</td>
<td>6.01</td>
</tr>
<tr>
<td>ARDRegression+</td>
<td>0.847</td>
<td>3.86</td>
<td>5.90</td>
</tr>
<tr>
<td>HuberRegressor</td>
<td>0.803</td>
<td>4.33</td>
<td>6.90</td>
</tr>
<tr>
<td>HuberRegressor+</td>
<td>0.803</td>
<td>4.14</td>
<td>6.78</td>
</tr>
<tr>
<td>Lars</td>
<td>0.861</td>
<td>3.88</td>
<td>5.86</td>
</tr>
<tr>
<td>Lars+</td>
<td>0.861</td>
<td>3.66</td>
<td>5.76</td>
</tr>
<tr>
<td>LassoLars</td>
<td>0.856</td>
<td>3.84</td>
<td>5.81</td>
</tr>
<tr>
<td>LassoLars+</td>
<td>0.856</td>
<td>3.66</td>
<td>5.73</td>
</tr>
<tr>
<td>PassiveAggressiveRegressor</td>
<td>0.529</td>
<td>7.20</td>
<td>9.35</td>
</tr>
<tr>
<td>PassiveAggressiveRegressor+</td>
<td>0.529</td>
<td>6.38</td>
<td>8.69</td>
</tr>
</tbody>
</table>

As shown in Table 4, based on fitting function generated by linear models such as Ridge Regression and Lasso Regression, this paper compares the errors between original predictive value and new predictive value. The experimental results show that, new predictive values proposed in this paper can reduce predictive error of all linear models to a certain extent. This is because the purpose of all linear models is to generate a fitting function, that represents the relationship between feature vector and predictive value. However, there is always an error between predictive value and true value. Based on fitting function of the linear model, this paper proposes a method for computing predictive value. The new predictive value can be closer to the true value than original predictive value, so predictive errors of all linear models can be improved to a certain extent.

6. Conclusion

The work of this paper includes following two parts. Firstly, based on fitting function of linear model, a change interval of predictive value is defined. If the relationship between true value and interval can be correctly judged, the prediction error can be effectively reduced. Secondly, when predicting the test data, the relationship between true value and interval can not be accurately judged. Therefore, according to distribution of training data in change interval, the relationship between true value and change interval of the predictive
value is estimated. According to the relationship, new predictive value is computed to replace the original predictive value. Tested on standard data sets, new predictive values can reduce the predictive error to a certain extent.

The paper estimates the relationship between true value and change interval based on probability, so it can only improve predictive error to a certain extent. If the relationship between true value and change interval can be computed concretely, the improvement of predictive error will be more obvious. In the next step, this paper will continue to study how to analyze the relationship between true value and change interval concretely based on classification algorithm.

References


