Extreme Learning Machine based on Capped ℓ1 Regularization and Pinball Loss Function

Zihao Zhou
College of Scinence
Xi'an Shiyou University
Shaanxi, China

Jinge Li
College of Computer Science
Xi'an Shiyou University
Shaanxi, China

Jiayi Jie
College of Chemistry and Chemical Engineering
Xi'an Shiyou University
Shaanxi, China

Zhenyu Zhu
College of Scinence
Xi'an Shiyou University
Shaanxi, China

Xibo Ming
College of Scinence
Xi'an Shiyou University
Shaanxi, China

ABSTRACT
Extreme Learning Machines (ELM) traditionally employ the squared loss function as the training criterion. However, this function is highly sensitive to outliers, which can amplify their impact on training outcomes of the model, causing the model to deviate from the true data distribution and reducing robustness of the model. Additionally, traditional ELM may encounter overfitting issues when dealing with high-dimensional dataset. To tackle these issues, this study introduces an innovative ELM frameworkthat integrates capped ℓ1 regularization with pinball loss function, termed as Cℓ1-PELM. The capped ℓ1 regularization helps prevent overfitting, and the pinball loss function, due to its linear relationship with the error, effectively mitigates the adverse effects of outliers on model training. This paper employs an iterative reweighting algorithm to optimize the objective function, ensuring rapid convergence of the model during the training process. Experimental results on 18 real-world datasets demonstrate that Cℓ1-PELM exhibits superior robustness, generalization performance, and stability in comparison to other advanced algorithms, particularly in environments with outliers.

General Terms
Machine Learning, Neural Network, Algorithms, Modeling

Keywords
Extreme learning machine; Capped ℓ1 regularization; Pinball loss function; Robustness

1. INTRODUCTION
ELM [1-2] generate random connection weights for the input and hidden layers, along with the biases for the neurons in the hidden layer, which remain unchanged during the training process. By determining the number of hidden layer neurons, the output weights can be calculated by applying the Moore-Penrose inverse matrix method to obtain the unique optimal solution. Compared with traditional training algorithms, ELM stands out for its simple implementation, quick training, and excellent generalization performance. It has gained significant attention and found application in multiple areas, including face recognition [3], electricity market forecasting [4], and fuel cell system [5]. However, outliers in the datasets may lead to overfitting of the model. ELM typically uses the squared loss function, which performs best when the errors are normally distributed, but the actual error distribution may not meet this assumption. To minimize the influence of outliers on model accuracy while maintaining the training speed and generalization advantages of ELM, constructing a robust ELM model in the field of machine learning is particularly necessary and significant.

Currently, improvements to robustness of ELM are primarily focused on two aspects. Firstly, the optimization of the loss function. Wang et al. [6]introduced the pinball loss function into ELM to minimize quantile errors, effectively suppressing the impact of outliers on the decision function. Yang et al. [7]applied the maximum mixed correlation entropy criterion to semi-supervised ELM,enhancing ability of the modelto resist outliers in the dataset, thereby improving robustnessthe model.Secondly, the introduction of regularization terms in ELM is aimed at preventing overfitting. Bala et al. [8]introduced an online sequential ELM based on ℓ21 norm regularization, suitable for real-time data processing, reducing the time and memory consumption required for the model retraining each time new data is added.Dai et al.[9] proposed a novel regularization method to address the needs of multi-dimensional output tasks in ELM. This method extends the traditional ℓ1/2 regularization to ℓ2,1/2 regularization. Through this extension, it is possible to enhance robustness of the model to outliers while maintaining model performance.

This paper presents an innovative ELM model that integrates capped ℓ1 regularization and pinball loss function, solved through an iterative reweighting method, with the aim of enhancing robustness of the model in the face of outliers in the dataset. The organization of the paper is as follows: the second section offers a concise overview of the theoretical foundations of ELM; the third section outlines the proposed algorithm; the fourth section offers experimental descriptions and result analysis; finally, summarizes the paper.

2. RELATED WORKS
2.1 Capped ℓ1 Regularization
Capped ℓ1 regularization [10,11] is an improved method over the traditional ℓ1 regularization[12], designed to more effectively approximate the ℓ0 regularization[13].The mathematical formulation of capped ℓ1 regularization is as follows:

$$\delta_u(\beta_j) = \min(u, |\beta_j|) \qquad (1)$$

Where u is threshold and $u > 0$ .As depicted in Figure 1, when the absolute value of the output weight is less than the threshold u , capped ℓ1 regularization behaves identically to ℓ1 regularization; when the absolute value of the output weight exceeds the threshold u , the capped ℓ1 regularization is capped at the threshold u . This approach effectively balances complexity of the modeland its robustness against outliers.

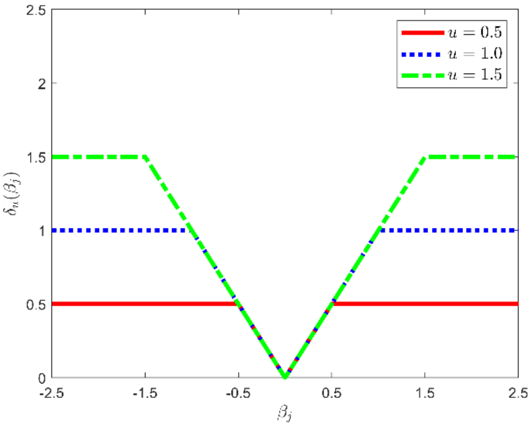


Figure 1: Capped ℓ1 Regularization

2.2 Pinball Loss Function

Pinball loss function [14] is the loss function designed for quantile regression, which can effectively reduce sensitivity of the model to outliers. It is defined as follows:

$$l_p(e_i) = \begin{cases} pe_i, & e_i \geq 0, \\ (p-1)e_i, & e_i < 0 \end{cases} \quad (2)$$

Where p denotes the quantile. By adjusting the value of p , it is possible to estimate different quantiles.As depicted in Figure 2, the pinball loss function exhibits linearity and asymmetry. When the error e_i is greater than 0, the growth rate of the squared loss function increases more rapidly than that of the pinball loss function with the increase in error; the converse is also true.

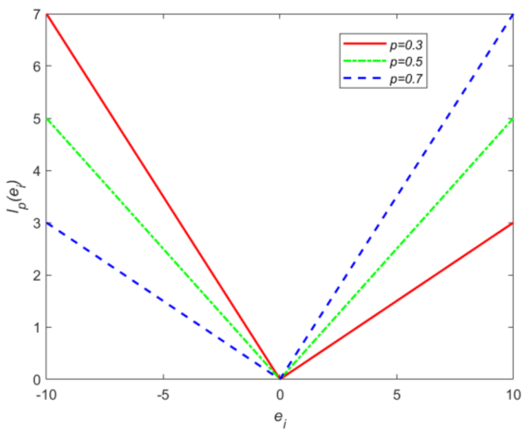


Figure2: Pinball loss function

2.3 Iterative Reweighted Robust Regularization ELM

To tackle the sensitivity of the squared loss function to outliers, this paper employs a reweighting approach to incorporate the L1 loss function into ELM to enhance the robustness of the model.This model can be realized by addressing the following optimization problem:

$$\begin{aligned} \min_{\beta} \quad & \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{N}\|e_i\|_1 \\ \text{s.t.} \quad & h(x_i)\beta = y_i - e_i, \quad i = 1,\ldots,N \end{aligned} \quad (3)$$

Where C represents the regularization parameter, which is utilized to control the complexity of themodel. Lagrangian function for equation (3) is constructed as follows:

$$L_{\ell_2}(\beta, e_i, \alpha_i) = \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{N}\|e_i\|_1 - \sum_{i=1}^{N}\alpha_i(h(x_i)\beta - y_i + e_i) \quad (4)$$

Where α_i is Lagrange multiplier. This paper employs an iterative reweighting method to solve equation (4). Each iteration is akin to tackling a weighted least squares problem, where the weights are determined by the errors from the previous iteration. Utilize the Karush-Kuhn-Tucker (KKT) conditions to obtain the best solution:

$$\begin{cases} \dfrac{\partial L_{\ell_2}(\beta_j, e_i, \alpha_i)}{\partial \beta_j} = 0 \Rightarrow \beta_j = h(x_i)^T\alpha_i \\ \dfrac{\partial L_{\ell_2}(\beta, e_i, \alpha_i)}{\partial e_i} = 0 \Rightarrow \alpha_i = Cw(e_i)e_i \\ \dfrac{\partial L_{\ell_2}(\beta, e_i, \alpha_i)}{\partial \alpha_i} = 0 \Rightarrow h(x_i)\beta - y_i + e_i = 0 \end{cases} \quad (5)$$

Where $w(\cdot)$ represents the weight function for ℓ1 norm, specifically $w(e_i) = 1/\max(|e_i|, 10^{-6})$. From equation (5), the output weights can be derived as follows:

$$\beta = \begin{cases} \left(\dfrac{I}{C} + H^T W_N H\right)^{-1} H^T W_N Y, & N \geq L, \\ H^T\left(\dfrac{I}{C} + W_N H H^T\right)^{-1} W_N Y, & N < L. \end{cases} \quad (6)$$

Where $H = [h(x_1), \ldots, h(x_N)]^T$
$W_N = diag\{w(e_1), \ldots, w(e_N)\}$.

3. ELM WITH CAPPED ℓ1 REGULARIZATION AND PINBALL LOSS FUNCTION

To improve the generalization performance of ELM, this paper proposes Cℓ1-PELM. The optimization problem of it is formulated as follows:

$$\begin{aligned} \min_{\beta, e_i} \quad & \sum_{j=1}^{L}\delta_u(\beta_j) + C\sum_{i=1}^{N} l_p(e_i) \\ \text{s.t.} \quad & h(x_i)\beta = y_i - e_i, i = 1,\ldots,N \end{aligned} \quad (7)$$

The optimal solution is obtained by constructing the Lagrangian function and employing KKT conditions.

$$\beta = \begin{cases} \left(\dfrac{W_L}{C} + H^T W_N H\right)^{-1} H^T W_N Y & N \geq L \\ W_L^{-1} H^T\left(W_N H W_L^{-1} H^T + \dfrac{I}{C}\right)^{-1} W_N Y & N < L \end{cases} \quad (8)$$

Where W_L represents the weight function for capped ℓ1 regularization, specifically

$$W_L = diag\{w_L(\beta_1), w_L(\beta_2), ..., w_L(\beta_L)\} \quad (9)$$

$$w_L(\beta_j) = \begin{cases} 0, & |\beta_j| \geq u, \\ 1/\max\{|\beta_j|, 10^{-6}\}, & |\beta_j| < u. \end{cases} \quad (10)$$

And W_N represents the weight function for pinball loss function, specifically

$$W_N = diag\{w_N(e_1), w_N(e_2), ..., w_N(e_N)\} \quad (11)$$

$$w_N(e_i) = \begin{cases} p/\max(e_i, 10^{-6}), & e_i \geq 0, \\ (p-1)/e_i, & e_i < 0. \end{cases} \quad (12)$$

The algorithm of Cℓ1-PELM is as follows:

Input: $\{(x_i, y_i)\}_{i=1}^N, L, C, p, u, t_{\max}, \mathrm{H}, \eta;$

Step 1: Initialize $W_N^{(0)} = \mathrm{I}; W_L^{(0)} = I; t = 1;$

Step 2: Calculate the output weightsusing Equations (6).

Step 3: Compute errors $e_i = h(x_i)\beta^{(t)} - y_i;$

Step 4: Update W_L and W_N using Equations (10) and (12).

Step 5: Update the output weights $\beta^{(t+1)}$ using Equation(8).

Step 6: If $t > t_{max}$ or $\| \beta^{(t+1)} - \beta^{(t)} \| < \eta$, stop the iteration and obtain the solution $\beta = \beta^{(t+1)}$. Otherwise $\beta^{(t)} = \beta^{(t+1)}$, return to Step 3.

Output: The weights β.

4. EXPERIMENTS

this study compares Cℓ1-PELM with the regularized extreme learning machine (RELM) [15], weighted regularized extreme learning machine (WELM) [16], iteratively reweighted robust regularized extreme learning machine (ℓ1ELM, ℓ1- ℓ1ELM) [17], and pinball loss-based extreme learning machine (PELM) [6] to validate the effectiveness of it. Parameters C, p, u are picked from sets (2^-19 2^-18 2^-17…2^17 2^18 2^19), (0.05 0.1 0.15…0.9 0.95 1), (0.0001 0.001 0.01 0.1 1 10 100 1000 10000),respectively. η is set to 0.001. RMSE is used to assess the performance of each model. All algorithms employ the Sigmoid function as the activation function, with a maximum iteration limit of 20. For each parameter configuration, the experiments are conducted independently 10 times to obtain the average test RMSE. The optimal parameters are determined based on the lowest average RMSE. The values of all datasets are normalized to the range (-1,1). The experimental hardware platform features an Intel Core i7-12700 processor running at 4.70 GHz, 16 GB of RAM, and a 64-bit Windows 10 OS. The programming environment is MATLAB R2020b.

4.1 Robustness Experimental Analysis

To verify the robustness of Cℓ1-PELM in handling dataset containing outliers, this paper employed a random sampling method to generate outliers, and randomly sampled values between the minimum and maximum outputs of the training samples and added these values to the designated outputs to create outliers [17].Table 1 presents detailed information of 18 real-world datasets, including the dimensionality of the datasets, the sizes of the training and test sets. Experiments were conducted on datasets with outlier ratios of 0%, 10%, and 20%, respectively. Table 2 presents the experimental findings, with the best results emphasized in bold.

As shown in Table 2, Cℓ1-PELM achieved the lowest RMSE on 11 datasets under the scenario with no outliers (0%). When the outlier ratio increased to 10%, Cℓ1-PELM attained the optimal prediction accuracy on 14 datasets. Even when the outlier ratio further increased to 20%, Cℓ1-PELM maintained the best prediction accuracy on 16 datasets. It is not difficult to observe that Cℓ1-PELM is suitable for datasets containing outliers, and as the proportion of outliers increases, the robustness and stability of the model improve.

Table 1. Information on 18 real-world datasets

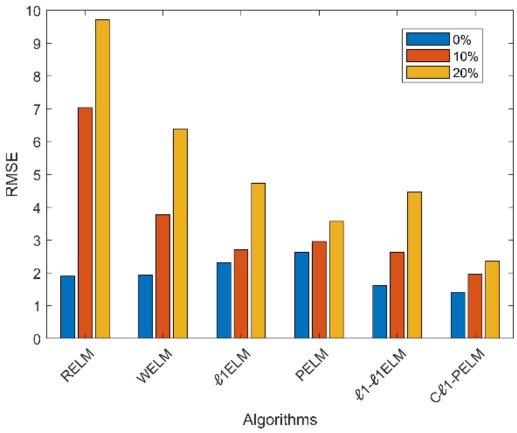
datasets	Feature dimension	Training set	Testing set
Machine	6	140	69
Diabetes	2	20	23
Cooling	8	400	368
Mpg	7	200	192
Yacht	6	200	108
BH	13	300	206
Concrete	8	600	430
NO2	7	300	200
Pollution	15	40	20
Pyrim	27	40	34
ENBC	8	400	368
Triazines	60	120	66
Servo	4	120	47
Bodyfat	14	160	92
Abalone	7	2000	2177
Airfoil	5	1000	503
MG	6	700	685
Space_ga	6	1800	1307

It is noteworthy that, compared to the other five algorithms, the RMSE of Cℓ1-PELM exhibits only minimal fluctuation with the increase in the ratio of outliers, as shown in Figure 3 (a). RELM uses the squared loss function demonstrates the poorest performance on datasets containing outliers. WELM, which employs aniterativereweighting method, experiences a significant increase in RMSE as the ratio of outliers grows. ℓ1ELM that utilizes an ℓ1 loss functionshows better performance due to ability of the ℓ1 norm to somewhat mitigate the impact of outliers. ℓ1- ℓ1ELM, which uses the ℓ1 norm for both the loss function and regularization term, further enhances robustness of ℓ1- ℓ1ELM. The PELM, which employs the pinball loss function, can adapt to datasets with outliers by adjusting the quantile parameter p, thus demonstrating relatively good performance. However, when compared with the Cℓ1-PELM proposed in this paper, these algorithms all exhibit greater fluctuations in response to changes in the ratio of outliers, and their RMSE values are not as optimal as those of the Cℓ1-PELM.The phenomenon in Figure 3 (a) is the same as that in Figure3 (a).

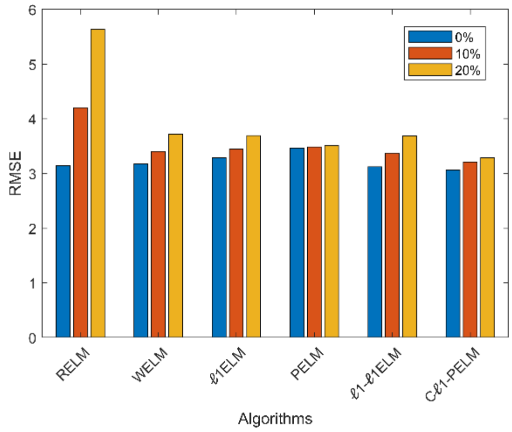
Table 2 Comparison of RMSE experimental results on real-world datasets

datasets	Outlier ratio	RELM	WELM	ℓ1ELM	PELM	ℓ1- ℓ1ELM	Cℓ1-PELM
Diabetes	0%	0.6171	0.6185	0.6244	0.6202	0.6008	**0.5902**
	10%	0.6874	0.6463	0.6481	0.6444	0.6112	**0.6096**
	20%	0.7227	0.6445	0.6535	0.6328	0.6163	**0.6134**
Mpg	0%	2.7911	2.7873	2.8082	2.8007	2.7939	**2.7825**
	10%	4.0918	**2.8401**	2.8815	2.8789	2.8577	2.8427
	20%	5.3817	3.2517	3.0269	2.9305	3.0336	**2.9167**
Yacht	0%	1.9021	1.9360	2.3088	2.6243	1.6115	**1.4117**
	10%	7.0368	3.7690	2.7096	2.9632	2.6337	**1.9675**
	20%	9.7171	6.3775	4.7369	3.5731	4.4640	**2.3636**
Concrete	0%	**6.4257**	6.5276	6.7397	6.9040	6.7773	6.7361
	10%	9.2527	7.4908	7.5653	7.6599	7.5582	**7.3435**
	20%	11.9875	8.4785	8.4698	7.9129	8.5101	**7.6945**
Pollution	0%	41.2155	**40.0596**	41.9972	41.4652	41.4997	40.8425
	10%	48.5809	44.5537	45.1348	45.0374	43.6350	**43.5461**
	20%	53.6576	48.2616	47.3794	45.6842	45.5953	**44.3591**
BH	0%	**3.8803**	4.0380	4.0233	3.9963	4.0033	3.9814
	10%	4.9008	4.1636	4.1711	4.1548	4.1312	**4.1166**
	20%	6.4469	4.4858	4.3243	4.2542	4.3508	**4.1977**
NO2	0%	**0.5135**	0.5168	0.5226	0.5227	0.5223	0.5222
	10%	0.5895	**0.5235**	0.5334	0.5250	0.5330	0.5248
	20%	0.7242	0.5466	0.5552	0.5252	0.5568	**0.5230**
Pyrim	0%	0.0864	0.0833	0.0818	0.0808	0.0802	**0.0789**
	10%	0.0990	0.0816	0.0828	0.0828	**0.0811**	0.0819
	20%	0.1095	0.0936	0.0893	0.0832	0.0817	**0.0805**
ENBC	0%	**1.4003**	1.4071	1.4875	1.5554	1.4322	1.4233
	10%	3.3833	1.8603	1.6957	1.6987	1.6358	**1.5908**
	20%	4.9116	2.6779	2.2271	1.7866	2.2397	**1.6901**
Machine	0%	57.3521	55.8167	55.2525	55.2559	56.4308	**53.3809**
	10%	93.8634	62.7115	57.1769	56.8396	59.378	**54.5161**
	20%	126.0512	67.4567	59.4492	58.1149	63.3377	**56.7770**
Triazines	0%	0.1350	0.1362	0.1339	0.1330	0.1328	**0.1324**
	10%	0.1408	0.1382	0.1396	0.1345	0.1379	**0.1335**
	20%	0.1464	0.1454	0.1444	0.1365	0.1410	**0.1343**
Servo	0%	0.5942	**0.5712**	0.6004	0.5931	0.6079	0.5944
	10%	0.9802	0.7562	0.6805	0.6598	0.6893	**0.6062**
	20%	1.1891	0.8974	0.7565	0.6633	0.7565	**0.6285**
Bodyfat	0%	0.0034	**0.0028**	**0.0028**	**0.0028**	0.0029	**0.0028**
	10%	0.0074	0.0033	**0.0029**	**0.0029**	0.0030	**0.0029**
	20%	0.0137	0.0047	0.0031	**0.0029**	0.0033	0.0030
Abalone	0%	2.1793	2.1900	2.1964	2.1765	2.1926	**2.1722**
	10%	2.6630	2.1743	2.1759	2.1792	2.1716	**2.1722**
	20%	3.6169	2.2320	2.2044	2.1820	2.2018	**2.1782**
MG	0%	**0.2265**	**0.2265**	0.2267	**0.2265**	0.2266	**0.2265**
	10%	**0.2265**	0.2266	0.2272	**0.2265**	0.2279	0.2266
	20%	**0.2264**	0.2268	0.2270	0.2265	0.2269	0.2266
Cooling	0%	**2.1686**	2.2153	2.3265	2.3738	2.2673	2.2431
	10%	3.7425	2.5159	2.5011	2.5151	2.4196	**2.3354**
	20%	5.1184	3.2732	2.9698	2.6084	2.9362	**2.4921**

Airfoil	0%	3.1448	3.1784	3.2834	3.4680	3.1225	**3.0628**
	10%	4.2038	3.4038	3.4498	3.4878	3.3746	**3.2130**
	20%	5.6325	3.7226	3.6950	3.5089	3.6851	**3.2896**
Space_ga	0%	**0.2000**	**0.2000**	**0.2000**	**0.2000**	0.2001	**0.2000**
	10%	0.2583	0.2010	0.2201	0.2010	0.2055	**0.2000**
	20%	0.3673	0.2077	0.2386	0.2019	0.2170	**0.2000**



(a) Yacht dataset



(b) Airfoil dataset

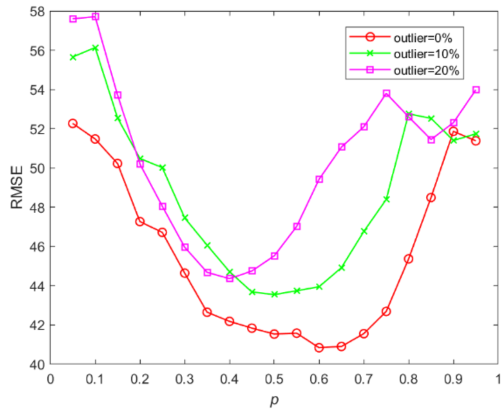
Figure3: The RMSE of RELM, WELM, ℓ1ELM, PELM, ℓ1- ℓ1ELMand Cℓ1-PELM algorithms under (a) Yacht, (b) Airfoil datasets with 0%, 10%, and 20%levels of outliers

4.2 Parameter Analysis

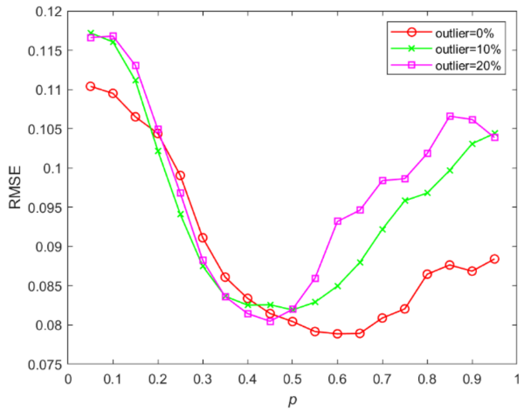
4.2.1 Parameter p Analysis

In the model proposed in this paper, the pinball loss function offers the advantage of mitigating the adverse effects of outliers by setting different quantile parameters p. To identify the optimal parameter p, this study employed a grid search approach across each real dataset. The objective of this method is to train the model to achieve the most effective output weights across each realdatasets, thereby enhancing the predictive accuracy and robustness of model. By doing so, we ensure that the model remains efficient and stable even when confronted with challenging data characteristics.

To investigate the impact of p on the performance of Cℓ1-PELM, we conducted experimental on Pollution and Pyrim datasets under varying outlier ratios,as illustrated in Figure 4, the experimental outcomes are displayed. It is evident that for Pollution dataset, the optimal p values corresponding to the lowest RMSE were 0.6, 0.5, and 0.4 at outlier ratios of 0%, 10%, and 20%, respectively. For Pyrim dataset, the optimal p values were 0.65, 0.5, and 0.45 at outlier ratios of 0%, 10%, and 20%. This finding indicates that Cℓ1-PELM can effectively control the rate of error growth through p tuning. Moreover, the results reveal that when datasets contain a higher ratioof outliers, selecting smaller parameter values is a more appropriate.



(a) Pollution dataset



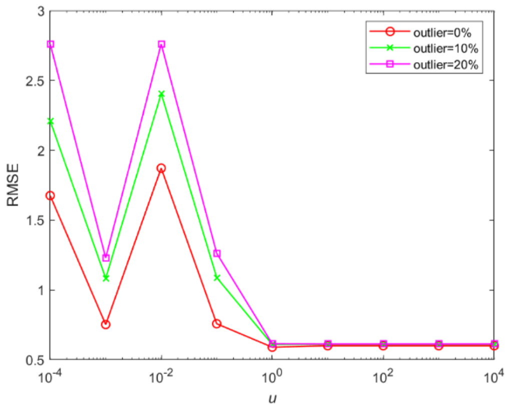
(b) Pyrim dataset

Figure.4 Influence of parameters on RMSE

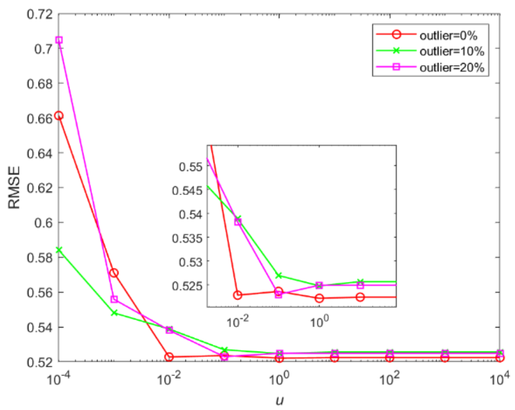
4.2.2 Parameter u Analysis

This paper employs the cappedℓ1 norm as a regularization term for the model. The capped ℓ1 regularization term includes a threshold parameter u, which allows for the restriction of output weights to a fixed size u when they exceed the threshold parameter u. This effectively limits the excessive increase of output weights, thereby preventing model overfitting. During the experimental evaluation on real datasets, a grid search method was utilized to determine the optimal parameter u . By employing this approach, the predictive performance and robustness of modelare enhanced.

To investigate the impact of u on the performance of Cℓ1-PELM, we selected Diabetes and NO2 datasets and conducted experiments under varying outlier ratios. The results of the experiments are depicted in Figure 5. Figure 5(a) illustrates RMSE trends of Cℓ1-PELM on Diabetes dataset at different outlier ratios. It can be observed that as the parameter u increases, the RMSE initially decreases, then increases, and finally decreases again to stabilize, reaching its minimum value at $u=100$ and remaining constant thereafter. Figure 5(b) shows RMSE trends on NO2 dataset with three different outlier ratios. The results indicate that RMSE decreases to a minimum value and then stabilizes as u increases. Overall, under different outlier ratios, RMSE of Cℓ1-PELM can be effectively regulated by appropriately selecting u , thereby enhancing predictionaccuracyof the model.



(a)Diabetes dataset



(b)NO2 dataset

Figure.5 Influence of parameters on RMSE

5. CONCLUSION

This paper proposes a novelELM regression model that integrates capped ℓ1 regularization and pinball loss function. Compared with traditional regularized ELM models, Capped ℓ1 regularization mitigates overfitting by limiting the magnitude of u , thereby reducing the model complexity.To address the issue of outliers in datasets, the model employs pinball loss function with different quantile parameters to reduce the adverse effects of outliers on its performance. The model is solved using an iterative reweighting algorithm to accelerate the model training speed. Experiments were conducted on 18 real-world datasets with varying ratios of outliers to validate robustness of the proposed Cℓ1-PELM and compare it with RELM, WELM, PELM, ℓ1ELM, and ℓ1-ℓ1ELM. The results demonstrate the Cℓ1-PELM outperformed other algorithms in terms of robustness and stability in most cases.

In future research, the model proposed in this paper holds broad application potential. Particularly within the context of naive bayes models, the introduction of the pinball loss function is expected to further enhance classification accuracy. This improvement will be especially beneficial when dealing with imbalanced datasets, as it will optimize model performance more effectively.

6. DATA AVAILABILITY

No new datasets were generated in the course of this study. All datasets utilized in this research are available from the authors upon request.

7. ACKNOWLEDGMENTS

The work was supported by the College Students' Innovation and Entrepreneurship Training Programof Xi'an Shiyou University under Grant No.S202310705098.

8. CONFLICT OF INTEREST

The authors have no conflict of interest. All co-authors have seen and agree with the contents of the manuscript, and there is no financial interest to report.

9. REFERENCES

[1] Huang G B, Zhu Q Y, Siew C K. 2004Extreme learning machine: a new learning scheme of feedforward neural networks[C]// IEEE international joint conference on neural networks.

[2] Wang J, Lu S, Wang S H, et al.2022 A review on extreme learning machine[J]. Multimedia Tools and Applications.

[3] Lal P V, Srilakshmi U, Venkateswarlu D. 2021Face recognition system using extreme learning machines[C]// 5th International Conference on Information Systems and Computer Networks (ISCON).

[4] Loizidis S, Kyprianou A, Georghiou G E. 2024Electricity market price forecasting using ELM and Bootstrap analysis: A case study of the German and Finnish Day-Ahead markets[J]. Applied Energy.

[5] Quan R, Liang W, Wang J, et al. 2024An enhanced fault diagnosis method for fuel cell system using a kernel extreme learning machine optimized with improved sparrow search algorithm[J]. International Journal of Hydrogen Energy.

[6] Wang K, Ding X. 2019Pinball loss based extreme learning machines[C]//IOP Conference Series: Materials Science and Engineering.

[7] Yang J, Cao J, Xue A. 2020Robust maximum mixture correntropy criterion-based semi-supervised ELM with variable center[J]. IEEE Transactions on Circuits and Systems II: Express Briefs.

[8] Bala R, Dagar A, et al. 2021A novel online sequential extreme learning machine with L2, 1-norm regularization for prediction problems[J]. Applied Intelligence.

[9] Dai Y, Zhang Y, Wu Q.2024 A pruning extreme learning machine with L 2, 1/2 regularization for multi-dimensional output problems[J]. International Journal of Machine Learning and Cybernetics.

[10] Nie F, Huo Z, Huang H. 2017 Joint capped norms minimization for robust matrix recovery[C]//The 26th International Joint Conference on Artificial Intelligence.

[11] Ma J, Yang L, Sun Q. 2020Capped L1-norm distance metric-based fast robust twin bounded support vector machine[J]. Neurocomputing.

[12] Balasundaram S, Gupta D. 20141-Norm extreme learning machine for regression and multiclass classification using Newton method[J]. Neurocomputing.

[13] Wang X, Wang K, She Y, et al. 2023Zero-norm ELM with non-convex quadratic loss function for sparse and robust regression[J]. Neural Processing Letters.

[14] Anand P, Rastogi R, Bharti A. 2018A Pinball loss based Extreme Learning Machine for pattern classification[C]IEEE Symposium Series on Computational Intelligence.

[15] Martínez-Martínez J M, 2011Escandell-Montero P, Soria-Olivas E, et al. Regularized extreme learning machine for regression problems[J]. Neurocomputing.

[16] Deng W, Zheng Q, Chen L. 2009 Regularized extreme learning machine[C] IEEE symposium on computational intelligence and data mining.

[17] Chen K, Lv Q, Lu Y, 2017et al. Robust regularized extreme learning machine for regression using iteratively reweighted least squares[J]. Neurocomputing.