

# Extreme Learning Machines Based on Least Absolute Deviation and Their Applications in Analysis Hard Rate of Licorice Seeds

Liming Yang<sup>1,2</sup>, Junjian Bai<sup>1</sup> and Qun Sun<sup>3</sup>

**Abstract:** Extreme learning machine (ELM) has demonstrated great potential in machine learning and data mining fields owing to its simplicity, rapidity and good generalization performance. In this work, a general framework for ELM regression is first investigated based on least absolute deviation (LAD) estimation (called LADELIM), and then we develop two regularized LADELIM formulations with the  $l_2$ -norm and  $l_1$ -norm regularization, respectively. Moreover, the proposed models are posed as simple linear programming or quadratic programming problems. Furthermore, the proposed models are used directly to analyze the hard rate of licorice seeds using near-infrared spectroscopy data. Experimental results on eight different spectral regions show the feasibility and effectiveness of the proposed models.

**Keywords:** Extreme learning machine, robust regression, least absolute deviation estimation, near-infrared spectroscopy

## 1 Introduction

Extreme learning machine (ELM) [Huang, Siew, and Zhu (2006)] is a popular and important learning algorithm for single-hidden-layer feedforward neural networks (SLFNs) [Huynh and Won (2008); Tiago, Francisco, Rui, and Carlos (2014); Huang, Song, and You (2015)], and has been successfully applied in regression and classification problems. Compared with traditional neural networks, the main advantages of ELM are that it runs fast and is easy to implement. Its hidden nodes and input weights are randomly generated and the output weights are expressed analytically. ELM overcomes some drawbacks of traditional neural networks, such as local minima, imprecise learning rates and slow convergence rates [Deng, Zheng, and Chen (2009); Zhang and Luo (2015); Zhua, Miao, and Qing (2014)].

---

<sup>1</sup> College of Science, China Agricultural University, Beijing, 100083, China.

<sup>2</sup> Corresponding author: Liming Yang; E-mail: cauyanglm@163.com

<sup>3</sup> College of Agriculture and Biotechnology, China Agricultural University, Beijing, 100193, China

ELM outputs its weights based on the least-squares estimation (LSE) [Xiang, Nie, Meng, Pan, and Zhang (2012)], but LSE is known to be highly sensitive to outliers. Therefore the traditional ELM is sensitive to noise and outliers, which may lead to poor performance. To tackle this drawback, the robustness of ELM has been investigated recently. For example, Huynh and Won proposed a weighted ELM [Huynh and Won (2008)], Deng et al. presented a regularized ELM [Deng, Zheng, and Chen (2009)], Zhang et al. developed an approximation algorithm [Zhang and Luo (2015)], and Horata et al. proposed robust ELM formulations to reduce the effects of outliers [Horata, Chiewchanwattana, and Sunat (2013); Zhua, Miao, and Qing (2014)]. These researches show that ELM is obviously affected by outliers, and thus reducing the effect of outliers is extremely important and necessary to build a robust ELM.

Least-absolute-deviation (LAD) estimation [Cao and Liu (2009); Yang, Liu, and You (2011)] is more resistant to outliers than least-squares estimation. The regression coefficients of LAD are estimated by minimizing the sum of the absolute values of the residuals, which involves nonsmooth optimization problem. Thus there are relatively few researches in LAD.

Inspired by LAD strategy, we present in this investigation a robust ELM regression framework based on LAD estimation. Two regularized LADELm formulations are developed. Furthermore, the proposed models are reformulated as convex programming with low computational burden.

Near-infrared (NIR) spectroscopy [Wang, Xue, and Sun (2012); Yang, Gao, and Sun (2015)] is based on the absorption of electromagnetic radiation ranging from 4000 to 12000 $cm^{-1}$ . NIR spectra can provide rich information on molecular structure. Recently, NIR spectroscopy has demonstrated great potential in the analysis of complex samples owing to its simplicity, rapidity and nondestructivity, and has been successfully applied to analyze the chemical ingredients or quality parameters of compounds.

Licorice is a traditional Chinese herbal medicine. Its seed is characterized by hardness. Usually, the hard rate of licorice-seed is determined by soaking the seeds, although this method is time-consuming and sometimes destroys the seeds. Therefore, developing a fast and nondestructive analysis technique for determining hard rate of licorice seeds is important and could promote the application of hard seeds in cultivation.

The main contributions of this work are summarized as follows:

- We first present an ELM framework based LAD regression (called LADELm) and then develop two regularized LADELm formulations with  $l_2$ -norm and  $l_1$ -norm regularization respectively.

- We reformulate the proposed models into as simple linear programming or quadratic programming problems with global solutions.
- The proposed models can yield robust solutions and thus are more resistant to outliers.
- The proposed methods are directly applied to analyze the hard rate of licorice seeds using NIR spectroscopy data.

Throughout the paper we adopt the following notations. An arbitrary dimension vector of ones is denoted by  $e$  and  $|\cdot|$  denotes absolute value operator.

The rest of the paper is organized as follows: Section 2 briefly summarizes ELM. Section 3 proposes a robust ELM framework based on LAD. Section 4 further develop two regularized LADELm formulations. Experiments for licorice-seed hard rate are analyzed in Section 5. Finally, we conclude this work and provide some ideas for future research in Section 6.

## 2 Background

### 2.1 Extreme learning machine (ELM)

ELM is a type of SLFNs and has been successfully applied to both classification and regression problems. Here, we give a brief definition of ELM regression; a more detailed description of ELM is available in the literature [Huang, Siew, and Zhu (2006); Jose, Martinez, and Pablo (2011)].

The essence of ELM is that its hidden-layer parameters are not necessarily tuned, and training error is minimized. Specifically, given a set of  $N$  patterns:  $\{(x_i, t_i), x_i \in R^n, t_i \in R, i = 1, \dots, N\}$ , where  $x_i$  is the input vector, and  $t_i$  is the target value. The goal of regression problem is to find a relationship between  $x_i$  and  $t_i$ , ( $i = 1, \dots, N$ ). We expect to find a standard SLFN with  $L$  hidden nodes to approximate these  $N$  patterns with zero error, which means that the desired output for the  $j$ -th pattern is

$$\sum_{i=1}^L \beta_i g(w_i \cdot x_j + b_i) = t_j, j = 1, \dots, N \quad (1)$$

where  $w_i$  is the weight vector connecting the  $i$ -th hidden node with the input node, and  $b_i$  denotes the bias term of the  $i$ -th hidden node. The  $\beta_i$  is the output weight from the  $i$ -th hidden node to the output node. The  $g(x)$  is an activation function and  $g(w_i \cdot x_j + b_i)$  is the output of the  $i$ -th hidden node. The linear system (1) is equivalent to the following matrix equation

$$H\beta = T \quad (2)$$

with

$$\mathbf{H} = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_L \cdot x_1 + b_L) \\ \vdots & \dots & \vdots \\ g(w_1 \cdot x_N + b_1) & \dots & g(w_L \cdot x_N + b_L) \end{bmatrix}_{N \times L}, \mathbf{T} = \begin{bmatrix} t_1 \\ \vdots \\ t_N \end{bmatrix}$$

where  $\beta = (\beta_1, \beta_2, \dots, \beta_L)^T$ .  $H$  is defined as the hidden layer output matrix, the  $i$ -th column of which is the  $i$ -th hidden node output with respect to the input  $x_i$ . The  $T$  is the desired output.

Huang et al. pointed out that the input weights  $w_i$  and hidden layer biases  $b_i$  for the SLFN are not necessarily tuned during training and may be assigned values randomly. Based on this scheme, Huang et al. proposed a simple SLFN algorithm, called ELM, the goal of which is to find a least-squares solution of the linear system (2). This can be posed as the following optimization

$$\min_{\beta} \|H\beta - T\|_2^2 \quad (3)$$

which is a normal quadratical programming with no constraints. With  $H^T H$  being positive definite, its optimal solution  $\hat{\beta}$  can be obtained by

$$\hat{\beta} = \mathbf{H}^\dagger \mathbf{T} \quad \text{where} \quad \mathbf{H}^\dagger = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \quad (4)$$

where  $H^\dagger$  is the Moore-Penrose generalized inverse of matrix  $H$ . The regression function of ELM is given by

$$f(x) = \sum_{i=1}^L \hat{\beta}_i g(w_i \cdot x + b_i) \quad (5)$$

## 2.2 Regularized extreme learning machine

To prevent overfitting data, the regularized ELM were investigated in literature [Deng, Zhang, and Chen (2009); Jose, Martinez, and Pablo (2011)]. A type version of the regularized ELM models is

$$\min_{\beta} \mu \|\beta\|_2^2 + \|T - H\beta\|_2^2 \quad (6)$$

where  $\mu > 0$  is a tuning parameter that balances  $l_2$ -norm regularization and empirical error. The problem (6) is also a convex optimization and thus it has global optimal solution. When  $\mu = 0$ , it is equivalent to the original ELM. Thus, model (6) is an expansion of the traditional ELM (3).

### 3 Extreme learning machine based on LAD method (LADELM)

Least-squares estimation (LSE) is suitable for the situations where the distribution of residual error assumes as a zero-mean Gaussian distribution. However in practical applications, this assumption is unrealistic. Least absolute deviations (LAD) method is more resistant to outliers and noise than LSE. LAD estimation, as a robust alternative to LSE, can overcome the shortcoming of LSE. Thus we replace the  $l_2$ -norm with the  $l_1$ -norm in ELM regression (3) to obtain a robust ELM based on LAD, called LADELM:

$$\min_{\beta} \|H\beta - T\|_1 \tag{7}$$

This involves a nonsmooth problem, and thus it is generally difficult to optimize. In the following section, we reformulate this problem as a linear programming.

Specifically, by introducing an additional variable  $z \in R^n$  with component  $z_j$  ( $j = 1, \dots, N$ ) such that the absolute values of the components for vector  $H\beta - T$  satisfy:

$$\left| \sum_{i=1}^L \beta_i g(w_i \cdot x_j + b_i) - t_j \right| \leq z_j, j = 1, \dots, N \tag{8}$$

then problem (7) can be reformulated as the following constrained optimization in variable  $(\beta, z)$

$$\begin{aligned} \min_{\beta, z} \quad & e^T z \\ \text{s. t.} \quad & \left| \sum_{i=1}^L \beta_i g(w_i \cdot x_j + b_i) - t_j \right| \leq z_j \\ & j = 1, \dots, N \end{aligned} \tag{9}$$

This is a linear programming and can be quickly solved. LADELM (7) can yield a robust solution since it is based on LAD strategy. Thus LADELM (7) is more resistant to outliers and noise.

Given a set of training samples  $\{(x_i, t_i), i = 1, \dots, N\}$ , the algorithm for LADELM (7) is summarized as follows:

**Algorithm 1**

- Choose a suitable hidden-layer units number  $L$  and an activation function  $g(x)$ . Assign randomly the hidden-unit parameters  $(w_i, b_i), i = 1, \dots, L$ .
- Calculate the hidden layer output matrix **H**.

- Solve the linear programming (9) to obtain the output weight  $\hat{\beta}$ , and then the  $\hat{\beta}$  is the optimal solution of LADELM (7).
- Use the model  $f(x) = \sum_{i=1}^L \hat{\beta}_i g(w_i \cdot x + b_i)$  to make regression prediction.

#### 4 Regularized extreme learning machine based on LAD method

In this section, we consider two regularized LADELM regression formulations with  $l_2$ -norm and  $l_1$ -norm regularization respectively.

##### 4.1 Regularized extreme learning machine based on LAD method

To prevent overfitting data, we add the  $l_2$ -norm regularization into the objective function of LADELM (7), and then obtain a regularized LADELM, called LADRELM

$$\min_{\beta} \|T - H\beta\|_1 + C\|\beta\|_2^2 \quad (10)$$

where  $C > 0$  is a tuning parameter that balances  $l_2$ -norm regularization and empirical error. Similarly, by introducing an additional variable  $p$  with components  $p_j$  ( $j = 1, \dots, N$ ) satisfying

$$\left| \sum_{i=1}^L \beta_i g(w_i \cdot x_j + b_i) - t_j \right| \leq p_j, j = 1, \dots, N \quad (11)$$

the problem (10) is posed as the following optimization in variable  $(\beta, p)$

$$\begin{aligned} \min_{\beta, p} \quad & e^T p + C\|\beta\|_2^2 \\ \text{s. t.} \quad & \left| \sum_{i=1}^L \beta_i g(w_i \cdot x_j + b_i) - t_j \right| \leq p_j \\ & j = 1, \dots, N \end{aligned} \quad (12)$$

This is a quadratic program problem with global optimal solution, and thus it is easy to solve. In this investigation, we only report the numerical simulation results for this problem. When  $C = 0$ , problem (10) is equivalent to the original LADELM (7). Thus RLADRELM (10) is an expansion of LADELM. Moreover, LADRELM is superior to the regularized ELM (6) by obtaining a robust solution.

##### 4.2 Sparse extreme learning machine based on LAD method

In this section, we consider a sparse LADELM model based on the  $l_1$ -norm regu-

larization. In particular, we incorporate the robust  $l_1$ -norm ( $\|\beta\|_1$ ) into the objective function for LADELM (7), and then get a sparse LADELM (called LADSELM)

$$\min_{\beta} \|H\beta - T\|_1 + \lambda \|\beta\|_1 \quad (13)$$

Where parameter  $\lambda > 0$  balances the empirical risk and the  $l_1$ -norm of the output vector  $\beta$ . Similarly, by introducing an additional variable  $q$  with components  $q_j$  satisfying:

$$\left| \sum_{i=1}^L \beta_i g(w_i \cdot x_j + b_i) - t_j \right| \leq q_j, j = 1, \dots, N \quad (14)$$

the problem (13) is reformulated as

$$\begin{aligned} \min_{\beta, q} \quad & e^T q + \lambda \|\beta\|_1 \\ \text{s. t.} \quad & \left| \sum_{i=1}^L \beta_i g(w_i \cdot x_j + b_i) - t_j \right| \leq q_j \\ & j = 1, \dots, N \end{aligned} \quad (15)$$

Furthermore, by adding a variable  $u$  with components  $u_j$  satisfying  $|\beta_i| \leq u_i$ , LADSELM (13) can be expressed as:

$$\begin{aligned} \min_{\beta, q, u} \quad & e^T q + \lambda e^T u \\ \text{s. t.} \quad & \left| \sum_{i=1}^L \beta_i g(w_i \cdot x_j + b_i) - t_j \right| \leq q_j, j = 1, \dots, N \\ & |\beta_i| \leq u_i, i = 1, \dots, L \end{aligned} \quad (16)$$

This is also a linear programming in variable  $(\beta, q, u)$  and can be quickly solved.

Note that LADSELM (13) is entirely different from the traditional regularized ELM (6) which is based on LSE and the  $l_2$ -norm regularization. The main benefits of the proposed LADSELM are summarized as follows:

- LADSELM is superior to the traditional ELM (3) because it is based on LAD. Thus it leads to a more robust solution and can reduce the sensitivity against outliers and noise.
- One appealing feature of  $l_1$ -norm regularization is that it can force sparsity. Thus LADSELM can yield a sparse model representation since it incorporates the  $l_1$ -norm regularization into its objective function. Therefore, LADSELM can control the tradeoff between the number of hidden-layer nodes and the empirical risk.

- LADSELM has low computational burden, only requiring to solve a linear programming.

Given a set of training samples  $\{(x_i, t_i), i = 1, \dots, N\}$ , according to the analysis above, the algorithm for solving LADSELM (13) summarized as follows:

### Algorithm 2

- Choose a suitable activation function  $g(x)$  and the number of hidden nodes  $L$ . Assign randomly hidden-layer parameters  $(w_i, b_i), i = 1, \dots, L$ .
- Calculate the hidden layer output matrix  $\mathbf{H}$ .
- Solve the linear programming (16) to obtain the output weight  $\hat{\beta}$ .
- Use the model  $f(x) = \sum_{i=1}^L \hat{\beta}_i g(w_i \cdot x + b_i)$  to make regression prediction.

## 5 Experiments

### 5.1 Sample set

The licorice seeds used in this experiment were harvested between 2002 and 2007, from various locations within China, including the Xinjiang municipality, Ningxia province, Inner-Mongolia municipality, Gansu province, Shanxi province and Heilongjiang province. The licorice-seed hard rate varied from 0.3% to 99.3%. After removing impurities, the seeds were put in a sample pool with the a diameter of 50mm. A total of 112 licorice seeds were used in the experiment.

The NIR spectra were acquired by using a spectrometer fitted with a diffuse reflectance fiber probe. Spectra were recorded over a range of 4000 to 12000 $cm^{-1}$  with a resolution of 8 $cm^{-1}$ . Each spectrum was the average of 32 scans. This procedure was repeated four times for each sample: twice from the front at different locations and twice from the rear at different locations. A final spectrum was taken as the mean spectrum of these four spectra. Consequently, the spectral data set contains 112 samples measured at 2100 wavelengths in the range of 4000 to 12000 $cm^{-1}$ . The NIR spectra of the licorice seeds are shown in Figure 1.

To evaluate the performance of the proposed models, numerical experiments are carried out on eight different spectral regions: 4000-6000 $cm^{-1}$ , 6000-8000 $cm^{-1}$ , 8000-10000 $cm^{-1}$ , 10000-12000 $cm^{-1}$ , 6000-10000 $cm^{-1}$ , 4000-8000 $cm^{-1}$ , 4000-10000 $cm^{-1}$  and 4000-12000 $cm^{-1}$ ; denoted A-H, respectively. Information on them is summarized in Table 1.



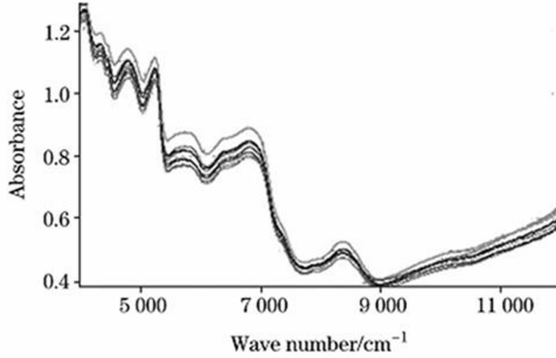


Figure 1: The near-infrared spectra of licorice samples.

Table 1: Near-infrared spectral sample regions of licorice seeds

Dataset	Spectral range( $cm^{-1}$ )	Number of samples	Number of wavelengths
region A	4000–6000	112	525
region B	6000–8000	112	525
region C	8000–10000	112	525
region D	10000–12000	112	525
region E	6000–10000	112	1050
region F	4000–8000	112	1050
region G	4000–10000	112	1575
region H	4000–12000	112	2100

### 5.2 Software and computing

We use MATLAB2012a to analyze the experiment results. The initial spectra were digitized by OPUS 5.5 software. After digitization, each spectrum in the 4000-12000 $cm^{-1}$  wavelength range was represented as a column vector; the length of the vector was defined by the number of wavelengths. The following toolboxes were used in this investigation:

MATLAB Statistics Toolbox.

MATLAB Linear Programming Toolbox.

MATLAB Quadratic Programming Toolbox.

### 5.3 Experimental design

To evaluate the proposed models, we specify the evaluation criteria before presenting the experiment results. Without loss of generality, let  $m$  be the number of test

samples;  $y_i$  denotes the real-value of sample  $x_i$ ;  $\hat{y}_i$  denotes the prediction value of  $x_i$  and  $\bar{y} = \frac{1}{m} \sum y_i$  is the average value of  $y_1, y_2, \dots, y_m$ . We use the following criteria to evaluate the algorithms [Peng (2010)]:

- SSE: the sum-squared error of test, defined as

$$SSE = \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (17)$$

SSE represents the fitting precision. Generally, the smaller the SSE, the better the estimation. However, an excessively small SSE value means that the regressor is probably overfitting.

- SST: the sum-squared deviation of test samples, defined as

$$SST = \sum_{i=1}^m (y_i - \bar{y})^2 \quad (18)$$

SST reflects the underlying variance of the test samples.

- SSR: the sum-squared deviation, defined as

$$SSR = \sum_{i=1}^m (\hat{y}_i - \bar{y})^2 \quad (19)$$

SSR reflects the explanation ability of the regressor. A larger SSR means that more statistical information was obtained from the test samples.

- SSE/SST: the ratio of the SSE to the SST of the test samples.
- SSR/SST: the ratio of the SSR to the SST of the test samples.

In general, a small SSE/SST means the estimates are consistent with the real values. Typically the SSR/SST increases as the SSE/SST decreases. In fact, an extremely small value for the SSE/SST is not desirable because it probably means that the regressor is overfitting. Therefore, a good estimator should strike the balance between the SSE/SST and SSR/SST.

- Average percent of selected hidden nodes numbers (PSHN).

In this investigation, we select sigmoid function

$$g(w, b, x) = \frac{1}{1 + \exp(-(w^T x + b))} \quad (20)$$

as the activation function in hidden layer. Moreover, we apply a tenfold cross-validation to the experiments.

The parameter  $\lambda$  in LADSELM implements a tradeoff between the empirical risk and the number of hidden-layer nodes. In general, when  $\lambda$  is large, the risk minimization is predominant leading to smaller regression error. Thus, the parameter  $\lambda$  should be optimized beforehand. The relationship between the SSE and parameter  $\lambda$  is illustrated in Fig. 2, from which we know that the SSE decreases when  $\lambda$  goes from 1 to 1000; at the same time, LADSELM produces smaller errors when  $\lambda > 1000$ . These findings help to select the value of  $\lambda$  in the following benchmark experiments.

In addition, the performance of the proposed LADELm and LADRELM also depends on the choice of the number of hidden nodes  $L$ . In this work, the value  $L$  is adjusted from the set  $\{20, 40, 60, 80, 100, 112, 500, 1000\}$  by a tenfold cross-validation. In each spectral region, the SSE is calculated and the optimal  $L$  is selected to minimize the SSE. The parameter  $L$  in the other models is set to be the same as LADELm.

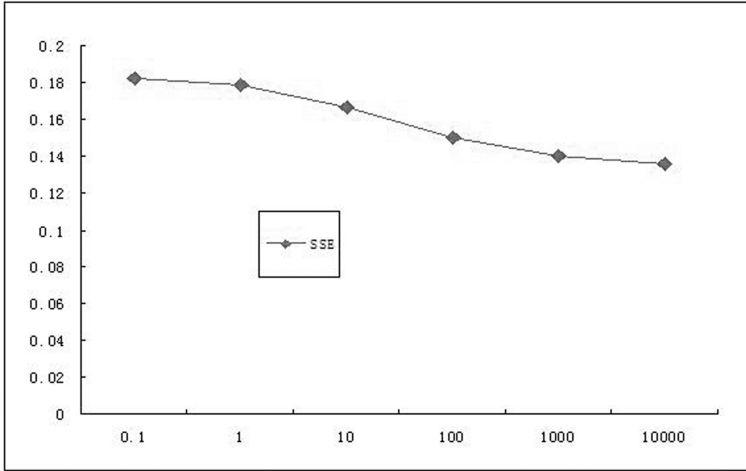


Figure 2: SSE versus parameter  $\lambda$  for LADSELM in region F.

For comprehensive evaluation, we compare the proposed models with the following baseline models: LSE, ELM, LAD, robust ELM (RELM) [Horata, Chiewchanwatana, and Sunat (2013)] and linear support vector regression (SVR) [Sangwook, RheeMan, and Minh (2015)] with parameter  $\varepsilon = 0.001$ . Experiment results on eight different spectral regions are reported in Table 2.

Table 2: Comparisons of eight algorithms: LSM, LAD, SVR, ELM, RELM, LADELm, LADRELM and LADSELM in linear case.

Regions	Methods	A	B	C	D	E	F	G	H
SSE (%)	LSM	35.34	45.00	83.54	164.93	40.34	31.27	32.67	52.49
	LAD	34.82	27.77	38.68	104.37	36.69	23.93	31.86	43.20
	SVR	33.77	34.36	75.30	99.13	40.38	32.55	43.52	41.55
	ELM	27.24	18.49	33.56	104.72	24.44	21.55	27.33	28.81
	RELM	18.55	18.84	31.03	103.61	21.69	19.19	22.11	23.43
	LADELm	14.33	17.08	21.05	79.77	19.23	16.26	17.75	25.84
	LADRELM	14.06	16.85	20.49	68.40	17.92	14.11	17.10	20.56
	LADSELM	13.87	16.56	19.57	60.12	17.99	13.57	16.79	20.23
SSE/SST (%)	LSM	12.49	12.55	24.02	49.15	14.10	10.86	11.59	17.05
	LAD	9.80	9.71	12.69	39.72	13.19	8.07	9.60	14.05
	SVR	10.23	11.94	27.08	38.57	12.94	10.65	16.19	12.76
	ELM	8.39	8.57	11.89	39.12	8.56	7.20	8.96	9.69
	RELM	5.95	7.26	9.89	33.15	7.03	6.38	8.29	8.50
	LADELm	4.87	6.91	8.55	31.00	6.11	6.19	6.73	8.34
	LADRELM	4.94	6.28	7.64	30.12	5.43	6.14	6.30	8.05
	LADSELM	4.88	6.11	7.13	26.15	5.72	6.11	5.41	7.89
SSR/SST (%)	LSM	70.23	90.63	79.33	68.25	91.15	77.61	87.68	71.97
	LAD	91.56	107.09	96.41	68.90	103.39	95.78	100.44	92.42
	SVR	80.97	82.78	82.00	65.20	86.06	93.02	76.51	80.38
	ELM	96.84	90.12	93.93	71.56	101.36	96.00	94.31	92.91
	RELM	94.01	91.41	91.01	93.00	93.00	96.62	92.61	91.40
	LADELm	95.02	94.10	92.44	70.99	93.88	94.11	93.06	92.68
	LADRELM	92.06	91.59	93.80	70.85	93.29	94.02	95.82	93.77
	LADSELM	93.70	94.15	92.64	71.92	93.57	93.86	93.67	92.33

To examine the sparseness of LADSELM, the average value of the PSHN is used as a measure of performance. We compare the performance of ELM, LADELm and LADSELM in the four broad spectral regions E, F, G, and H. Moreover, we choose two different values  $L$ ,  $L = \text{sample numbers}$  and  $L = \text{the number of wavelengths}$ . The average value of the PSHN on tenfold cross-validation is reported in Table 3.

In addition, we compare the proposed models with other models (LSE, LAD, SVR, ELM and RELM) with respect to the percent reduction in the SSE for LADELm and LADSELM. And the results are illustrated in Fig. 3 and Fig. 4 respectively.

Table 3: Comparisons of LAD, ELM and LADELm in terms of PSHN.

Regions	Methods	E	F	G	H
Hidden nodes numbers	ELM	100	100	100	100
$L =$ sample numbers	LADELm	80	69	70	56
	LADSELm	78	81	76	74
Hidden nodes numbers	ELM	88.76	80.86	76.38	70.90
$L =$ wavelength numbers	LADELm	79.05	64.00	56.00	61.00
	LADSELm	46.00	29.43	38.29	35.00

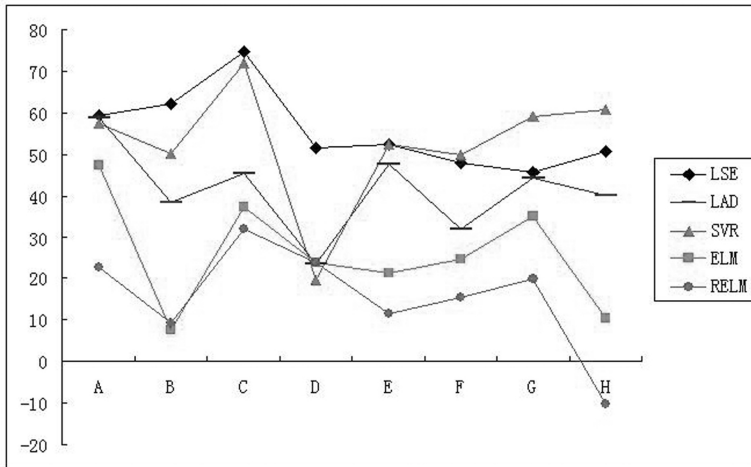


Figure 3: Percent reduction in the SSE of LADELm versus other models.

### 5.4 Experiment results

#### 5.4.1 Comparisons of the proposed models with other models

As expected, the above results show that the proposed models perform satisfactorily. The results are summarized as follows:

- Comparisons of the proposed models with the baseline models: LSE, LDA and SVR

Compared with the baseline models (LSE, LAD and SVR), the above results show that the proposed models (LADELm (7), LADRELm (10) and LADSELm (13)) achieve remarkable improvements in generalization for all eight spectral ranges.

Compared with LSE, the proposed LADELm, LADRELm and LADSELm

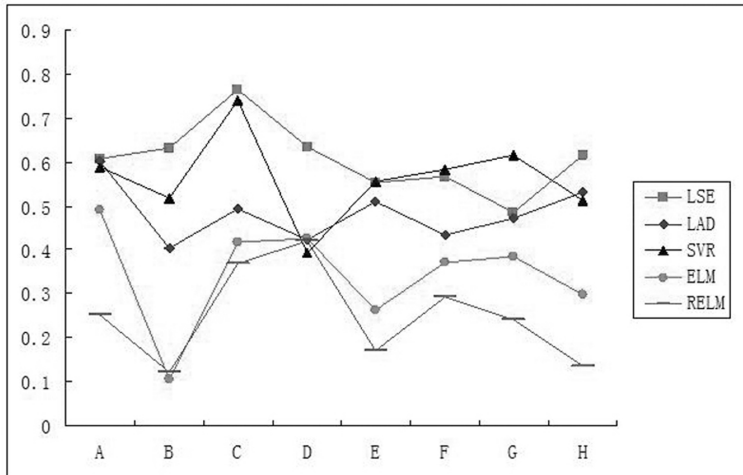


Figure 4: Percent reduction in the SSE of LADSELM versus other models.

decrease averagely the SSE by 55.59%, 59.48% and 60.77% , respectively. Especially, in the high-frequency region C, the values of the SSE decrease by 74.80%, 75.47% and 76.57%, respectively. At the same time, three models reduce the SSE/SST and SSR/SST in all eight spectral regions. Especially in region C, LADSELM and LADSELM decrease averagely the SSE/SST by 64.4%, 68.19% and 70.31%, respectively. Moreover the SSR/SST decreases more than 90% in seven of eight spectral regions.

Compared with SVR, the proposed LADELm and LADSELM decrease averagely the SSE by 52.74% and 56.32%, respectively. Meanwhile, the performance of LADSELM and LADSELM are significantly superior to SVR in terms of the SSE/SST and SSR/SST analysis.

Compared with LAD, the proposed LADELm, LADRELM and LADSELM reduce averagely the SSE by 41.33%, 46.42% and 48.39%, respectively. In terms of the SSE/SST, the proposed LADELm and LADSELM outperform clearly LAD in all eight spectral regions.

- Comparisons of the proposed models with the traditional ELM models

As expected, Table 2 illustrates that the proposed models achieves better performance than the traditional ELM (3) in all eight regions. To be more specific, compared with ELM, the proposed models, LADELm, LADRELM and LADSELM, decrease averagely the SSE by 25.92%, 32.27% and 34.48%, respectively. Moreover, three proposed models outperform clearly the ELM

(3) with respect to the SSR/SST. Especially in the two high-frequency regions C and D, the proposed LADSELM decreases the SSE/SST by 40.03% and 33.15%, respectively. At the same time, the three proposed model outperform ELM with respect to the SSR/SST in six of the eight regions.

- Comparisons of the proposed models with other robust ELM model

The performance of LADELM is superior to that of the robust ELM (RELM) [Horata, Chiewchanwattana, and Sunat (2013)] in seven of the eight spectral regions (the exception is region H). LADSELM yields smaller regression error than RELM, and decreases averagely SSE by 25.04% in eight regions. Compared with RELM in the two high-frequency regions C and D, LADSELM decreases the SSE by 36.93% and 41.97% respectively, and decreases the SSE/SST by 9.63% and 21.11% respectively. Finally, the performance of LADSELM is similar to that of RELM with respect to SSR/SST.

- Comparisons of ELM, LADELM, and LADSELM with respect to the PSHN.

In terms of the PSHN criterion, Table 3 shows that the proposed LADSELM (13) outperforms significantly the ELM (3) and LADELM (7), whereas

LADELM performs slightly better than ELM (3) in four spectral regions. These results suggest that LADSELM can select fewer hidden-node numbers than LADELM and ELM.

#### *5.4.2 Comparisons of the three proposed models*

The above results illustrate that the proposed LADSELM (13) and LADRELM (10) outperform LADELM (7) with respect to the SSE and SSE/SST in seven of eight regions, whereas there is no significant difference with respect to the SSR/SST. Compared with LADELM, two models LADSELM and LADRELM decrease averagely the SSE by 11.01% and 8.03% respectively. These results suggest that the regularized LADELM improves the generalization for analysis of the licorice-seed hard rate using NIR spectroscopic data. Moreover according to the above analysis, the effectiveness of the proposed methods can be arranged in the following order

$$\text{LADSELM} \geq \text{LADRELM} \geq \text{LADELM}$$

## **6 Conclusions and Future Directions**

We propose a robust ELM framework based on LAD regression. Moreover the proposed framework is used for the direct analysis the hard rate of licorice seeds using near-infrared spectroscopy data. We rigorously verify the proposed strategy

in different spectral regions. The main results of this work are summarized as follows:

- A robust ELM framework is first proposed based on LAD regression, called LADELm. Then we develop two regularized LADELm formulations with the  $l_2$ -norm and  $l_1$ -norm regularization, respectively.
- The proposed models are more resistant to outliers and noise than the original ELM which is based on LSE.
- Compared with the original ELM, the proposed LADRELM improves regression accuracy. The proposed LADSELM can yield a sparse solution due to incorporating  $l_1$ -norm regularization in its objective function.
- Three proposed ELM models can be reformulated as simple linear programming or quadratic programming problem, and thus are easy to solve.
- The proposed models are applied directly to analyze licorice-seed hard rate using near-infrared spectroscopy technology. Experiments results in different spectral regions show that the proposed methods obtain better generalization than traditional methods.

Analyzing licorice-seed hard rate is an important part of seed testing. The results presented herein show that it is possible to analyze licorice-seed hard rate by using the proposed models and near-infrared spectroscopy technology. We hope that this work will help further investigations of the hard rate of crop seeds and promote the application of hard seeds in cultivation. The question of building ELM classification framework based on LAD will be studied in the future work.

**Acknowledgement:** This work is supported by National Nature Science Foundation of China (11471010, 11271367).

## References

- Cao, H.; Liu, X.** (2009): System identification based on the least absolute criteria. *Techniques of Automation and Applications*, vol. 28, no. 7, pp. 8–10.
- Deng, W.; Zheng, Q.; Chen, L.** (2009): Regularized extreme learning machine. *in: IEEE Symposium on Computational Intelligence and Data Mining, CIDM'09*, pp. 389–395.
- Horata, P.; Chiewchanwattana, S.; Sunat, K.** (2013): Robust extreme learning machine. *Neurocomputing*, vol. 102, pp. 31–44.



- Huang, G. B.; Siew, C. K.; Zhu, Q. Y.** (2006): Extreme learning machine: Theory and applications. *Neurocomputing*, vol. 70, pp. 489–501.
- Huang, G.; Huang, G. B.; Song, S.; You, K.** (2015): You. trends in extreme learning machines: A review. *Neural Networks*, vol. 61, no. 1, pp. 32–48.
- Huynh, H. T.; Won, Y.** (2008): Weighted least squares scheme for reducing effects of outliers in regression based on extreme learning machine. *International Journal of Digital Content Technology and its Applications*, vol. 2, no. 3, pp. 40–46.
- Jose, M.; Martinez, M.; Pablo, E. M.** (2011): Regularized extreme learning machine for regression problems. *Neurocomputing*, vol. 74, pp. 3716–3721.
- Peng, X.** (2010): TSVR: An efficient twin support vector machine for regression. *Neural Networks*, vol. 23, no. 3, pp. 365–372.
- Sangwook, K.; ZH. Yu.; RheeMan, K.; Minho, L.** (2015): Deep learning of support vector machines with class probability output networks. *Neural Networks*, vol. 64, pp. 19–28.
- Tiago, M.; Francisco, S.; Rui, A.; Carlos, H. A.** (2014): Learning of a single-hidden layer feedforward neural network using an optimized extreme learning machine. *Neurocomputing*, vol. 129, pp. 428–436.
- Wang, Q.; Xue, W.; Sun, Q.** (2012): Quantitative analysis of seed purity for maize using near infrared spectroscopy. *Transaction of the Chinese Society of Agricultural Engineering*, vol. 28, pp. 259–264.
- Xiang, Sh. M.; Nie, F. P.; Meng, G. F.; Pan, Ch. H.; Zhang, Ch. G.** (2012): Discriminative least squares regression for multiclass classification and feature selection. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 11, pp. 1738–1754.
- Yang, L. M.; Go, Y. P., Sun, Q.** (2015): A new minimax probabilistic approach and its application in recognition the purity of hybrid seeds. *Computer Modeling in Engineering and Sciences*, vol. 104, no. 6, pp. 493–506.
- Yang, M.; Liu, Y.; You, Zh.** (2011): Estimating the fundamental matrix based on least absolute deviation. *Neurocomputing*, vol. 74, pp. 3638–3645.
- Yang, L. M.; Sun, Q.** (2012): Recognition of the hardness of licorice seeds using a semi-supervised learning method. *Chemometrics and Intelligent Laboratory Systems*, vol. 114, pp. 109–115.
- Zhang, K.; Luo, M.** (2015): Outlier-robust extreme learning machine for regression problems. *Neurocomputing*, vol. 151, pp. 1519–1527.
- Zhua, W.; Miao, J.; Qing, L.** (2014): Robust regression with extreme support vectors. *Pattern Recognition Letters*, vol. 45, pp. 205–210.

