

Model Optimization of Air Quality with M-ELM

Jianxiong Ye^{1, a} Wenzhen Zhou^{2, b} Zhigang Li^{3, c} Jinlan Zhou^{1, d}

¹Jiangxi Province Key Laboratory of precision drive & control, Nanchang Institute of Technology, China 330099

²School of Information Science and Engineering, Central South University, China 330013

³School of Mechanic Engineering, East China Jiaotong University, China 330013

^ajxlpjx@163.com, ^b315578522@qq.com, ^c214504207@qq.com, ^d214503760@qq.com

Keywords: Extreme learning machine, M-estimator, Model performance, Air quality model

Abstract. The extreme learning machine (ELM), which was originally proposed for "generalized" single-hidden layer feedforward neural networks (SLFNs), provides efficient unified learning solutions for the applications of clustering, regression, and classification. But when the training data have been contaminated, ELM can't guarantee the model accuracy. A novel hybrid way called M-ELM is proposed to adjust the output matrix of ELM model, this way combined M-estimator with ELM to reduce the noise influence. Experimental results on UCI (University of California at Irvine) datasets and air quality detection indicate that M-ELM performs competitively good, it can be used on design of air cleaner.

I Introduction

In recent years, due to its extremely fast training, and universal approximation capability, ELM has been becoming a significant research topic for machine learning. Single-layer ELM algorithms [1-4] and multi-layer ELM have been set up based on widespread SLFNs [5-11]. But when the samples are heavily polluted by noise, ELM can't ensure a good fitting accuracy, so M-estimation is used to solve this problem.

II Basic-ELM Introduction

2.1 ELM structure and its mathematical express

ELM is a feedforward networks, it includes only one input layer and a output layer, one or more hidden layers, Studies have proved that multi-layer network can learn any N distinct samples with any arbitrarily small error by no more than $2\sqrt{(m+2)N} (\ll N)$ hidden neurons, where m is the number of output neurons[4], this result means that the required number of hidden neurons of ELM can be significantly decreased in application; and the most popular form is single-hidden layer feedforward networks (SLFNs), the structure of SLFN can be described as figure 1.

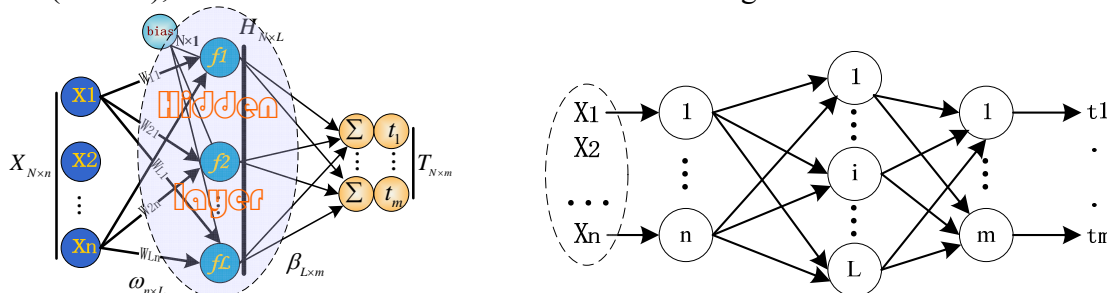


Fig. 1. Network structure constructed by ELM

As shown in figure 1, L means number of hidden neurons, g indicates activation function and m means the number of output neurons, $f(x) = \sum_{i=1}^L \beta_i g(a_i x_j + b_i) = t_j, j = 1, \dots, N$, it can be rewritten as $H\beta = T$, where

$$H(a_1, \dots, a_L, b_1, \dots, b_L, x_1, \dots, x_N) = \begin{bmatrix} g(a_1 \cdot x_1 + b_1) & \dots & g(a_L \cdot x_1 + b_L) \\ \dots & \dots & \dots \\ g(a_1 \cdot x_N + b_1) & \dots & g(a_L \cdot x_N + b_L) \end{bmatrix}_{N \times L}, \beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}, T = \begin{bmatrix} t_1^T \\ \vdots \\ t_L^T \end{bmatrix}_{N \times m}$$

$a_i = [w_{i1}, w_{i2}, \dots, w_{in}]$, so output matrix can be simply represented as follows:

$$\hat{\beta} = H^\dagger T \quad (1)$$

Where H^\dagger is the Moore-Penrose generalized inverse of H ^[2,15].

2.2 Main Problems in Learning

According to equation (1), how to calculate H^\dagger is important, and several ways have been proposed such as iterative method, orthogonalization method, orthogonal projection method and singular value decomposition(SVD)^[12], the problem lies in that if the training data is polluted, all these resolutions are meaningless. How to get a better output matrix needed further study.

III Combination of M-estimation and ELM

3.1 Working Process of M-ELM

The error is inevitable for sampling data, there are two kinds of them, one is small and random error, which is also called system error; another is gross error, it also call outliers, M-estimator is a kind of robust estimator and used to reduce the influence of this kind of error.

Consider regression model $T = H\beta + e$, a new method called ME-ELM is designed:

$$\hat{\beta} = (H^T P H)^{-1} H^T P T \quad (2)$$

Where P is the coefficient matrix, which should be tuned according to the training errors. For noise-free samples, all the coefficients of P are set as 1, it implies that ME-ELM works in the same way as ELM; if part of samples are moderately polluted, corresponding weights are less than 1, so as to weaken the poor influence; for those with gross errors, the relating weights tend to become much more smaller, so the bad influence can be greatly reduced. Generally speaking, adjusting matrix P plays an vital role in improving model accuracy.

The weights of P are regulated by estimation function ψ , which is derivation of static function

$\rho: \psi(x) = \frac{\partial \rho(x)}{\partial \beta}$, and optimization objective function $Q(\beta)$ can be defined as:

$$Q(\beta) = \sum_{i=1}^N \rho(e_i) = \sum_{i=1}^N \rho(T_i - H_i \beta_i) \quad (3)$$

$$\text{Solutions of (3) are called M-estimators: } \hat{\beta} = \arg \min_{\beta} \left(\sum_{i=1}^N \rho(T_i - H_i \beta_i) \right) \quad (4).$$

Many estimation functions ψ have been proposed, and most of them have similar results in improving model accuracy, Huber function is a popular one and is used in our research^[13].

$$\psi(x) = \begin{cases} x & |x| < k (k > 0) \\ k & |x| > k (k > 0) \end{cases}, \quad \rho(x) = \begin{cases} x^2 / 2 & |x| \leq k \\ k|x| - k^2 / 2 & |x| > k \end{cases}, \quad k=1.345 \quad (5)$$

3.2 Experimental Verification

The performance of M-ELM are tested with datasets shown in table 1, these datasets are all from UCI^[14-17]. Segmentation dataset is used to predict classification of images according to 19 attributes, the Iris is for species prediction of iris from four physical sizes, the Abalone is to estimate the age of abalone from 8 physical measurements; Bike Share dataset is used to determine the relation between bike rented hour and the data of weather, season and environment etc; CCPP contains 9568 data points taken from various sensors from 2006 to 2011 and is used to predict the energy output of electrical plant. All theses datasets are different in types, sizes and complexities, and for all the problems here, both the training data and testing data are chosen out randomly for each trial.

Table 1 Specification of benchmarking datasets

Datasets	Type Classification/Regression	#Train	#Test	#Attributes
Segmentation	C	110	100	19
Iris	C	80	70	5
Abalone	C/R	2000	2177	8
Bike Sharing (hours only)	R	10000	7397	13
CCPP	R	23920	23920	5

The relationship between the number of hidden neurons and network performance, as well as the working efficiency of M-ELM are studied. results are shown in table 3 when N=30, 60 and 100. It can be seen that testing accuracy improves with the increasing of hidden neurons number, and the training time increases synchronously.

Table 2 Testing Result of Benchmarking Datasets

Datasets	N=20			N=50		N=100	
	Hidden Neurons	Best RMSE for Test	Train Time	Best RMSE for Test	Total Train Time	Best RMSE for Test	Total Train Time
Iris	40	0.1286	0.3588	0.0571	1.0140	0.0386	2.2620
Segmentation	40	0.1300	0.1872	0.1100	1.3104	0.1000	0.9516
Abalone	40	0.0743	0.4368	0.0731	1.2948	0.0735	2.4648
fried	400	0.0423	152.9278	0.0365	368.9112	0.0407	742.8144
Bike Sharing	400	0.0149	72.7745	0.0142	166.2971	0.0138	330.0981
CCPP	400	0.0503	76.4873	0.0508	224.6570	0.0499	272.8145

IV Conclusions

Normal ELM works on the minimal least square error(LSE) principle, it can approximate N arbitrary samples with zero error, but if the data are contaminated, the network constructed by ELM and other algorithms can't work well, comparison tests manifest that M-ELM is useful in this case, it can weaken the outliers influence and achieve a better output matrix of SLFN. Now we are using it in air quality prediction, the samples from different sensors are often contaminated by random dust, so it is meaningful to put M-ELM in this situation.

V Acknowledgements

This work is partially founded by Science Project of Jiangxi Education Department (GJJ151129), Open Project Program of Jiangxi province key laboratory of precision drive & control, (PLPDC-KFKT-201625), Natural Science Foundation of China (CN) (51665016), Natural Science Foundation of Jiangxi Province (CN) (20151BAB207047).

References

- [1] Guangbin Huang, Qinyu Zhu, Cheekheong Si. Extreme learning machine:Theory and applications[J]. Neuro computing, 70,2006:489–501
- [2] Guangbin Huang, Hongming Zhou, Xiaojian Ding, Rui Zhang. Extreme Learning Machine for Regression and Multiclass Classification[J]. IEEE Trans. on Systems, Man, and Cybernetics-Part B: Cybernetics, (42)2, 2012:513-529
- [3] Nanying Liang, Guangbin Huang. A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks[J]. IEEE Trans. on Neural Networks, (17)6,2006:1411-1424

- [4] G.B. Huang. Learning capability and storage capacity of two-hidden layer feedforward networks[J]. IEEE Trans. Neural Networks, (14)2,2003:274–281
- [5] G.B. Huang, L. Chen, and C.-K. Siew. Universal approximation using incremental constructive feedforward networks with random hidden nodes[J]. IEEE Trans. Neural Networks, (17)4,2006:879–892
- [6] G.-B. Huang and L. Chen. Convex incremental extreme learning machine[J]. Neurocomputing, (70)16–18,2007:3056–3062
- [7] G.B. Huang, L. Chen. Enhanced random search based incremental extreme learning machine[J]. Neurocomputing, (71)16–18,2008:3460–3468
- [8] Miche Y, Sorjamaa A, Bas P, Simula O, Jutten C, Lendasse A (2010) Optimal pruned extreme learning machine. IEEE Trans Neural Netw (21)1,2010:158–162
- [9] Yimin Yang, Q.M. Jonathan Wu. Two-Layer extreme learning machine for dimension reduction[C]. Proceedings of ELM-2015, (2):31-42
- [10] J. Tang, C. Deng, and G.-B. Huang. Extreme learning machine for multilayer perceptron[J]. IEEE Trans. Neural Netw. Learn. Syst., to be published.
- [11] H. Zhou, G.B. Huang, Z. Lin, H. Wang, Y. C. Soh. Stacked extreme learning machines[J]. IEEE Trans. Cybern., (45)9,2015:2013–2025
- [12] C.R. Rao, S.K. Mitra. Generalized Inverse of Matrices and its Applications[M].Wiley, New York, 1971
- [13] James O. Street, Raymond J. Carroll, David Ruppert. A Note on Computing Robust Regression Estimates Via Iteratively Reweighted Least Squares[J]. The American Statistician,(42)2,1988:152-154
- [14] Lichman, M. (2013). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science
- [15] Fanaee-T, Hadi, and Gama, Joao. Event labeling combining ensemble detectors and background knowledge[J]. Progress in Artificial Intelligence, (2)2,2014: 113-127
- [16] Pınar Tüfekci. Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods[J]. International Journal of Electrical Power & Energy Systems, (60)9, 2014: 126-140
- [17] <http://archive.ics.uci.edu/ml/datasets/Abalone>