2014 IEEE International Conference on Systems, Man, and Cybernetics
October 5-8, 2014, San Diego, CA, USA

DYNAMICALLY REGULARIZED HARMONY LEARNING OF GAUSSIAN MIXTURES

Hongyan Wang, Jinwen Ma

De a men f In f ma i n Science, Sc l f Ma ema ical Science
e ing Uni e i , ei ing, 100871, C ina

ABSTRACT

In i a e, a d namicall eg la i ed a m n lea ning D alg i m i ed f a ian mi e lean-ing i a fa i e fea e f b ada i e m del elec-i n and c n i en a ame e e ima i n. S eci call , n-de e f ame f a e ian ing- ang a m -n lea ning e ili e e a e age S ann n en f e e i babili e am le a a eg la i a i n e m be-ing c n lled b a cale fac e a m n f nc i n n a ian mi e inc ea ing f m 0 1 d namicall . I i dem n a ed b e e imen n b n e ic and eal- ld da a e a e D alg i m can n l elec e c ec n mbe f ac al a ian in e da a e, b al b ain e ma im m li eli d M e ima f e a ame e n e ac al mi e. M e e , e D al-g i m i calable and can be im lemen ed n a big da a e .

Keywords a ian mi e m del elec i n eg la-i a i n ma im m li eli d.

1. INTRODUCTION

A a e ible and e f l a i ical l f da a anal - i and in f ma i n ce ing, ni e mi e m del 1 a f nd i a lica i n in man blem , c a cl e ing anal i, image egmen a i n and eec ec gni i n. A-m ng e e a lica i n , a ian mi e a e idel ed and e e al a i ical lea ning me d a e been ed deal i i ind f m del , c a e EM alg i-m 2 and e me d f m men 3 . e all a-med a e n mbe f a ian cl e in a da a e i e- n. e e, in man in ance i e inf -ma i n i n a ailable. en, e elec i n f an a i-a e n mbe f a ian , called m del elec i n, m be made i e a ame e lea ning in mi e . , e gene al a ian mi e m deling i ac all a c m nd m deling blem f b a ame e e ima i n and m d-el elec i n ic i a a e c m lica ed and dif c l a 4 .

e c n en i nal a f l ing i c m nd mi - e m deling blem i elec an imal n mbe k^* f

---

in en Ma, e c e nding a , el 8 -10- 27 0 09, Email ma ma . .ed .cn.

a ian a e cl e in e da a e ia ne f e in-f ma i n, c ding and a i ical elec i n c i e ia c a A ai e In f ma i n C i e i n 5 , a e ian In fe ence C i-e i n IC , Minim m De c i i n eng MD 7 , and Minim m Me age eng MM 8 . e e, e alida ing ce f i me d i c m a i nall e en-i e beca e eed e ea e en i e a ame e lea ning ce a a la ge n mbe f ible k.

Since e 1990 , me a i ical lea ning a ac e a ea ed l e i blem. Di ic le ce e 9 and e e ible m Ma c ain M n e Ca l MCMC 10 a e ical im lemen a i n f e ind fa ac -e ic e c a ic im la i n . e e c a ic im-la i n me d gene all elie n in en i e am ling and a e e ime-c n ming. e ec nd ind i e a e ian m del ea c ba ed n imi ing e a ia i nal b nd 11 . e i d i n e i ed lea ning n ni e mi e 12, 13 ic in d ce ce ain c m e i i e mec ani m in- e mi e m del c a e m del elec i n can be made ada i el d ing a ame e lea ning i a im li ed MM m del elec i n c i e i n ing e C mec a-ni m.

Al e na i el , e a e ian ing- ang a m -n lea ning 14, 15, 1 a al ided a ne a i ical lea ning mec ani m a ma e m del elec i n ada i el d ing a ame e lea ning. I a al ead been im lemen -ed n a ian mi e lea ning and e e al a m n lea ning alg i m a e al been e abli ed f a ian mi e 17, 18, 19 . Al g e a m n lea ning n e abili f ada i e m del elc i n, i a ame e e ima i n a a n able de ia i n f m e M e ima i n ic i c n i en i e a ame e .

In de l e i de ia i n blem, e c n ide e a m n lea ning a e M lea ning, i a eg-la i a i n e m f a e age n ega i e S ann n en f e e i babili e am le ing c n l e-cale c m le i f mi e m del . m i n f ie e e a d namicall eg la i ed a m n lea n-ing D alg i m f a ian mi e b adding a S ann n en eg la i a i n e m being c n lled b a cale fac e a m n lea ning. A e cale fac- i inc ea ed d namicall f m 0 1, ed algorithm transforms from the harmony learning with adaptive model selection to the conventional maximum likelihood learning so that the adaptive model selection and the ML estimation are both obtained at last.

The rest of the paper is organized as follows. We begin with a brief description of the BYY harmony learning system for Gaussian mixture in Section 2. Then, we present the derivation and analysis of the dynamically regularized harmony learning algorithm for Gaussian mixture in Section 3. Section 4 contains the experimental results on both synthetic and real-world data sets. Finally, we conclude briefly in Section 5.

2. BYY HARMONY LEARNING OF GAUSSIAN MIXTURES

The BYY harmony learning system describes each observation $x \in X \subset \Re^n$ and its corresponding inner representation $y \in Y \subset \Re^m$ via the two types of Bayesian decomposition of the joint density, $p(x,y) = p(x)p(y|x)$ and $q(x,y) = q(y)q(x|y)$, which are called Yang machine and Ying machine, respectively. Given a sample data set $D_x = \{x_t\}_{t=1}^N$ from the Yang or observable space, the aim of the BYY harmony learning system is trying to extract the hidden probabilistic structure of x with the help of y from specifying all aspects of $p(y|x)$, $p(x)$, $q(x|y)$ and $q(y)$ by maximizing the following harmony functional

$$H(p||q) = \int p(y|x)p(x)\ln[q(x|y)q(y)]dxdy. \quad (1)$$

If both $p(y|x)$ and $q(x|y)$ are parametric, the learning system is called a Bi-directional Architecture (BI-Architecture for short). Given a sample data set $D_x = \{x_t\}_{t=1}^N$, the BI-architecture of the BYY harmony learning system can be specified as follows. The inner representation y is discrete in $Y = \{1,2,\cdots,k\}$ (i.e., with $m=1$), while the observation x is continuous from a Gaussian mixture distribution. On the Ying space, we let $q(y=j) = \pi_j \geq 0$ with $\sum_{j=1}^k \pi_j = 1$. This is a prior probability distribution of Gaussian clusters for the mixture. On the Yang space, $p(x)$ is a latent probability density function (pdf) of a Gaussian mixture from which D_x are generated. Moreover, in the Ying path, $q(x|y=j) = q(x|m_j,\Sigma_j)$ is assumed to be a Gaussian density function with mean vector m_j and the covariance matrix Σ_j, while in the Yang path, $p(y=j|x)$ is constructed under the Bayesian principle by the following parametric form,

$$p(y=j|x) = \frac{\pi_j q(x|m_j,\Sigma_j)}{q(x|\Theta_k)}, \quad (2)$$

$$q(x|\Theta_k) = \sum_{j=1}^k \pi_j q(x|m_j,\Sigma_j), \quad (3)$$

where $\Theta_k = \{\pi_j, m_j, \Sigma_j\}_{j=1}^k$ and $q(x|\Theta_k)$ is just a Gaussian mixture model that will approximate the latent $p(x)$ via the harmony learning on the BYY learning system. Putting all these components into Eq.(1), we have

$$H(p||q) = E_{p(x)}\left[\sum_{j=1}^k h_j(X)\ln[\pi_j q(X|m_j,\Sigma_j)]\right], \quad (4)$$

where

$$h_j(X) = \frac{\pi_j q(X|m_j,\Sigma_j)}{\sum_{i=1}^k \pi_i q(X|m_j,\Sigma_j)}. \quad (5)$$

That is, $H(p||q)$ is the expectation of a function of the random variable X with respect to $p(x)$. With the sample data set D_x, we get an estimate of $H(p||q)$, called harmony function, as follows:

$$J(\Theta_k) = \frac{1}{N}\sum_{t=1}^N \sum_{j=1}^k h_j(x_t)\ln[\pi_j q(x_t|m_j,\Sigma_j)].$$

According to the theoretical and experimental results on this BI-architecture of the BYY harmony learning system for Gaussian mixture [20, 17, 18, 19], the maximization of $J(\Theta_k)$ is capable of making model selection adaptively during parameter learning when the actual Gaussian clusters are separated in a certain degree. That is, if we set k to be larger than the number k^* of actual Gaussian clusters in the sample data, the maximization of the harmony function can make k^* Gaussians match the actual ones and implicitly eliminate $k-k^*$ extra ones. However, as we mentioned previously, the original harmony learning suffers from inconsistent parameter estimation. So, we here introduce the regularization mechanism to transform the harmony learning into the ML learning such that adaptive model selection and consistent parameter estimation can be made simultaneously.

3. DYNAMICALLY REGULARIZED HARMONY LEARNING ALGORITHM

3.1. The Dynamic Regularization Mechanism

According to [21], $J(\Theta_k)$ can be divided into two terms,

$$J(\Theta_k) = L(\Theta_k) - O_N(p(y|x)), \quad (7)$$

where the first term is just the log-likelihood function, i.e.,

$$L(\Theta_k) = \frac{1}{N}\sum_{t=1}^N \ln\left(\sum_{j=1}^k (\pi_j q(x_t|m_j,\Sigma_j))\right), \quad (8)$$

while the second term is the average Shannon entropy of the posterior probability $p(y|x)$ over the sample data set $\mathcal{D} = \{x_t\}_{t=1}^N$,

$$O_N(p(y|x)) = -\frac{1}{N}\sum_{t=1}^N \sum_{j=1}^k p(j|x_t)\ln p(j|x_t). \quad (9)$$

Acc din g E . 7 , if $-O_N(p(y|x))$ i ie ed a a eg- la i a i e m, e a m , lea in g, i.e., ma imi - in g $J(\Theta_k)$, i a eg la i ed M lea in g ic a al ead been in e iga ed in 22, 23 b calin g e eg la i a i e m i a mall i i e mbe . w e e , ince e ee e eg la i a i cale c an a in e ca e f e a m , lea in g, e e in e iga i al ffe f m in c i en a ame e e ima i .

On e e an d, f m E . 7 e al a e

$$L(\Theta_k) = J(\Theta_k) + O_N(p(y|x)), \quad 10$$

w ic in dca e a e M lea in g i a eg la i ed a m lea in g i $O_N(p(y|x))$ a e eg la i a i e m. c l e eg la i a i ,a cale fac $\lambda(\geq 0)$ i in d ced,

$$L_\lambda(\Theta_k) = J(\Theta_k) + \lambda O_N(p(y|x)). \quad 11$$

If $\lambda = 0$, $L_\lambda(\Theta_k) = J(\Theta_k)$ i a m f c- i e i-ac i ec e f a ian mi e . If $\lambda = 1$, $L_\lambda(\Theta_k)$ i e l g-li eli d f c i f e a ian mi - e m del. a i i λ in c ea in g f m 0 1, ma imi - in g $L_\lambda(\Theta_k)$ c an ge f m e a m lea in g e M lea in g. e e c l e in c ea in g f λ a - ia el eali e ada i e m del elec i a e e i lea in g age an d e M e ima i a e al lea in g age.

3.2. The Fixed-point Learning Algorithm

A eac a e f ed amicall eg la i ed a m lea - in g i a eci c λ e c c a ed- in alg i m ma imi e $L_\lambda(\Theta_k)$.

c en ien ce e ili e e f ma e e en a i f π_j, i.e., $\pi_j = e^{\beta_j}/\sum_{i=1}^{K} e^{\beta_i}$, $j = 1, \cdots, k$, e e $\beta_j \in (-\infty, +\infty), j = 1, \cdots, k$. e in g e de i a i e f $L_\lambda(\Theta_k)$ i e ec β_j, m_j an d Σ_j, e ec i el be e e ge e f ll in g ed- in i e a i e lea in g alg i m

$$\hat{\pi}_j = \frac{\sum_{t=1}^{N} p(j|x_t)\gamma_j(t)}{\sum_{t=1}^{N}\sum_{i=1}^{k} p(i|x_t)\gamma_i(t)}; \quad 12$$

$$\hat{m}_j = \frac{\sum_{t=1}^{N} p(j|x_t)\gamma_j(t)x_t}{\sum_{t=1}^{N} p(j|x_t)\gamma_j(t)}; \quad 13$$

$$\hat{\Sigma}_j = \frac{\sum_{t=1}^{N} p(j|x_t)\gamma_j(t)(x_t-\hat{m}_j)(x_t-\hat{m}_j)^T}{\sum_{t=1}^{N} p(j|x_t)\gamma_j(t)}, \quad 14$$

w e e

$$\gamma_i(t) = 1 - \sum_{l=1}^{k}(p(l|x_t) - \delta_{il})\ln \pi_l p(x_t|m_l, \Sigma_l) + \lambda\sum_{l=1}^{k}(p(l|x_t) - \delta_{il})\ln p(l|x_t), \quad 15$$

w e e δ_{ij} i e ec e f c i .

In c m a i i e c en i al EM alg i m f a ian mi e 2 , ed ed- in lea in g al- g i m diffe l a e a gmen in g e m $\gamma_j(t)$. I can be ea il e i ed a en $\lambda = 1$, $\gamma_j(t) = 1$, e ed- in lea in g alg i m i e EM alg i m an d en $\lambda = 0$, e ed- in lea in g alg i m e e - igin al ed- in lea in g alg i m f ma imi in g e a m f c i $J(\Theta_k)$.

Ac all , $\gamma_j(t)$ im lemen a i al en ali ed c m e i- i e lea in g C mec an i m 24 - 25 a m del - elec i can be made ada i el d in g a ame e lea in g. A e ea l lea in g age, $\gamma_j(t) < 0$ ma a en . Acc d- in g E . 15 , e mean ec f j- a ian ill m e a a f m x_t. O e i e, if $\gamma_j(t) > 0$, e mean ec f e j- a ian ill be a ac ed x_t. S , f x_t, a - ian i $\gamma_j(t) > 0$ a e in e ile e e a ian i $\gamma_j(t) < 0$ a e l e .

w e e , e ed- in lea in g alg i m can g a- an ee e i i e de in i e e f eac c a ian ce ma i d in g e i e a i in ce $\gamma_j(t)$ ma be n ega i e. In de e c me i blem e e EM da e le f e c a ian ce ma i e , i.e., f cin g all $\gamma_j(t) = 1$ in E . 14 , in i eci c ca e. In fac , i im li ca i i a lica- ble an d ef cien in ce e c m e i i f ada i e m del elec i i main l am n g mean ec an d c lled b e mi in g i .

3.3. The Dynamic Evolution of λ

e f e di c e d amic e l i f λ i ime T d in g e lea in g ce . Acc din g eg la i a i mec an i m, λ ld be e mall an d in c ea e l l a e ea l lea in g age eali e ada i e m del elec i . en, a e e en age, λ can g l a a ic e eed an d e alg i m ill all c e ge a M l i . S , i i c cial c ec e e e ada i e m del elec- i a acc m li ed an d en c an ge lea in g age.

In de de ec e in g in e in d ce e S an n en f mi in g i in a ian mi - e , $H_\pi = -\sum_{j=1}^{k}\pi_j \ln \pi_j$. Ob i l , H_π i en i i e e c e f a mi e m del. If m del elec i i n c m le ed, e diffe en ce f H_π be een i e a i i c ide able. O e i e, e diffe en ce ld be e mall. i m i a e ad e ab l e c an ge a e f H_π be een i e a i , de n ed b

$$h_\pi(T) = \left|\frac{H_\pi(T) - H_\pi(T-1)}{H_\pi(T)}\right|, \quad 1$$

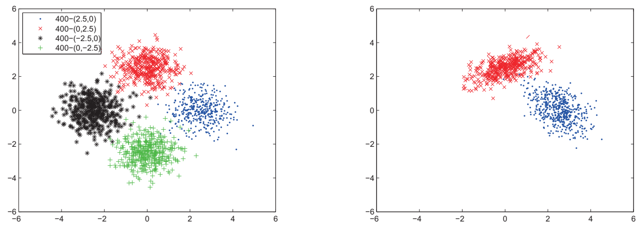
a an in dica f m del elec i . T i ime, i.e., e n m- be f i e a i e le lea in g ce i di ided in lea in g age acc din g a gi en e ld $\varepsilon_1(> 0)$ f i in dica . a i , if $h_\pi(T) > \varepsilon_1$, $\lambda(T)$ in c ea e a lw eed e w i e, i inc ea e a ig eed. Since $\lambda(T)$ i a med inc ea e e nen iall, i d namic e l i n ce i gi en a f llw ,

$$\lambda(T) = \begin{cases} \lambda_0 * \eta_1^T, & \text{if } h_\pi(T) > \varepsilon_1; \\ \lambda_0 * (\frac{\eta_1}{\eta_2})^{T^*} \eta_2^T, & \text{if } h_\pi(T) \leq \varepsilon_1, \end{cases} \quad 17$$

w e e λ_0 being a e mall i i e c n an i ini ial al e f λ, η_1, η_2 a e w i i e c n an w i c n an a $1 < \eta_1 < \eta_2$, and T^* i e n ing n c a $h_\pi(T^*) > h_0$ and $h_\pi(T^* + 1) \leq h_0$. en λ eac e l w e i n il e i e a i n c n e ge .

3.4. The Complete DRHL Algorithm

[redacted]



a

[redacted]

Table 1. e al e f e a ame e f e f ic da a e .

e da a e	a ian	m_i	σ_{11}^i	$\sigma_{12}^i(\sigma_{21}^i)$	σ_{22}^i	π_i	N_i
$\mathcal{S}_1$	1	2.50,0	0.50	0.00	0.50	0.25	400
(1 00)	2	0,2.50	0.50	0.00	0.50	0.25	400
	3	-2.50,0	0.50	0.00	0.50	0.25	400
	4	0,-2.50	0.50	0.00	0.50	0.25	400
$\mathcal{S}_2$	1	2.50,0	0.45	-0.25	0.55	0.34	544
(1 00)	2	0,2.50	0. 5	0.20	0.25	0.28	448
	3	-2.50,0	1.00	0.10	0.35	0.22	352
	4	0,-2.50	0.30	0.15	0.80	0.1	2 5
$\mathcal{S}_3$	1	2.50,0	0.10	-0.20	1.25	0.50	00
(1200)	2	0,2.50	1.25	0.35	0.15	0.30	3 0
	3	-1,-1	1.00	-0.80	0.75	0.20	240
$\mathcal{S}_4$	1	2.50,0	0.28	-0.20	0.32	0.34	8
(200)	2	0,2.50	0.34	0.20	0.22	0.28	5
	3	-2.50,0	0.50	0.04	0.12	0.22	44
	4	0,-2.50	0.10	0.05	0.50	0.1	32

g e ell a k^* a ian dem a ed b ei c line a e all ec g i ed and eac e ima ed a ian ma c e e ac al e acc a el .

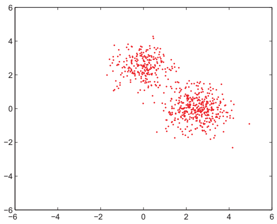


Table 2. e c m a i f e D and CEM^2 alg i m m del elec i and i ime.

Da a e	D		CEM^2	
	CMS e enc	ime	CMS e enc	ime
$\mathcal{S}_1$	100	52	84	11290
$\mathcal{S}_2$	100	85	5	1825
$\mathcal{S}_3$	100	145	72	4317
$\mathcal{S}_4$	9	4 0	5	554

Table 3. e c m a i f e D and CEM^2 alg i m a ame e e ima i acc ac .

Da a e	D	CEM^2
$\mathcal{S}_1$	0.0204	0.0204
$\mathcal{S}_2$	0.0171	0.0172
$\mathcal{S}_3$	0.03 3	0.03 3
$\mathcal{S}_4$	0.0308	0.0715

ig e an e f e D alg i m.

4.2. Unsupervised Classifications of Iris and Wine Data

e f e a l e D alg i m e e i ed cla i ca i f e I i and ine da a f m UCI Mac ine ea ing e i 28 . I i da a e c ain ee cla e , I i e ic l , I i i gin ica and I i Se a, and eac cla c i f 50 am le . Eac am le i 4-dimen i nal ec mea ing e lan m l g . In e e i-men n e I i da a e e e ini ial al e f k a 6 and e ini ial al e f e e a ame e a in im la-i n e e imen . ene all , e D alg i m ed a $k^* = 3$ i e imal cla i ca i acc ac 9 .7 . Onl e f m 150 am le a e mi cla i ed . e e , i i ible a e D alg i m c n e ge $k^* = 2$. Since e e a e I i b-cla e ic a e ngl e-la ed, me li e a e al acce $k^* = 2$.

e ine da a e i 13-dimen i nal and c n i f 178 am le f ee ine . In i ca e e e ce i da a e b e inci al c m nen anal i CA dimen-i n ed c i n ec n i e 29 and c e n l e ee inci le c m nen . e D alg i m i c nd c ed n e e ce ed da a i e ini ial al e f k a . E e imen al e l dem n a e a e D alg i m al a c n e ge i $k^* = 3$ and e acc ac f cla i-ca i can eac a 98.3 . Onl ee am le a e mi cla -i ed .

5. CONCLUSIONS

e a e in e iga ed e ela i i be een e a-m n lea ning and e M lea ning and b idged em -ing a eg la i a i e m e a e age S ann n en f e e i babili e am le. a ed n c a eg-la i a i mec ani m e c n c e d amicall eg-la i ed a m n lea ning D f a ian mi e . c n lling e cale fac f i eg la i a i e m d namicall inc ea e f m 0 1, e D alg i m - a f m e a m n lea ning i a ca abili f ada i e m del elec i n, and en g ad all an f m e c n en i nal ma im m li eli d lea ning b ain a c n i en a ame e e ima i n. M e e , e D al-g i m i calable and can be ed n a big da a e i ce ain da a mma i a i n ec ni e. E e imen al e l dem n a e a , n b n e ic and eal ld da a e , e D alg i m can n l elec e c ec n m-be f ac al a ian in a da a e , b al b ain e M e ima e f e a ame e in e mi e.

Acknowledgments.

i a ed b e a al Science nda i n f C ina f an 1171138 and 07710 1.

6. REFERENCES

1 Da id eel and Mac la n. ini e mi e m del , 2000.

2 ic a d A edne and me al e . Mi e den i ie , ma im m li eli d and e em alg i m. *SIAM Review*, 2 2 195 239, 1984.

3 eil E Da . E ima ing e c m nen f a mi e f n mal di ib i n . *Biometrika*, 5 3 4 3 474, 19 9.

4 A a igan. Di ib i blem in cl e ing. *Classification and Clustering*, age 45 72, 1977.

5 i g A ai e. A ne l a e a i ical m del iden i ca i n. *IEEE Transactions on Automatic Control*, 19 71 723, 1974.

ide n Sc a e al. E ima in g e dimen i n f a m del. *The Annals of Statistics*, 2 4 1 4 4, 1978.

7 ma i anen. M deling b e da a de c i - i n. *Automatica*, 14 5 4 5 471, 1978.

8 C i S. allace and Da id . D e. Minim m me - age len g and lm g c m le i . *The Computer Journal*, 42 4 270 283, 1999.

9 Mic ael D E c ba and Mi e e . a e ian den i- e ima i n and infe ence ing mi e . *Journal of the American Statistical Association*, 90 430 577 588, 1995.

10 S l ia ic a d n and e e een. On ba e ian anal i f mi e i an n n mbe f c m- nen i di c i n. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59 4 731 792, 1997.

11 C n an in C n an in l and A i idi i a . Un e i ed lea ning f ga ian mi e ba ed n a ia i nal c m nen li ing. *IEEE Transactions on Neural Networks*, 18 3 745 755, 2007.

12 Ma i A ig ei ed and Anil . ain. Un e- i ed lea ning f ni e mi e m del . *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 3 381 39 , 2002.

13 . M. C e ng. Ma im m eig ed li eli d i- a i al en ali ed em f den i mi e cl e ing i a ma ic m del elec i n. *IEEE Transactions on Knowledge and Data Engineering*, 17 705 7 1, 2005.

14 ei . ing- ang mac ine a ba e ian- llbac c eme f ni ed lea ning and ne e l n ec- an i a i n. In *Proceedings of the 1995 International Conference on Neural Information Processing (ICONIP?5)*, l me 2, age 977 988, 1995.

15 ei . e a m n , ni ed cl and a ma- ed m del elec i n f n e i ed and e i ed lea ning n ga ian mi e , ee-la e ne and me- bf- m m del . *International Journal of Neural Systems*, 11 01 43 9, 2001.

1 ei . a m n lea ning, c al cl, and l gical elf- gani ing n mi e m del . *Neural Networks*, 15 8 1125 1151, 2002.

17 in en Ma, ai n ang, and ei . A g adien b a m n lea ning ga ian mi e i a - ma ed m del elec i n. *Neurocomputing*, 5 481 487, 2004.

18 in en Ma and e ang. a m n lea ning n ni e mi e ada i e g adien im lemen a i n and an a ing cl mec ani m. *Neural Processing Letters*, 24 1 19 40, 200 .

19 in en Ma and efeng e. A fa ed- in b a m n lea ning alg i m n ga ian mi e i a ma ed m del elec i n. *Pattern Recognition Letters*, 29 701 711, 2008.

20 in en Ma. A ma ed m del elec i n am n - n i e mi e a e e ical anal i . In *International Joint Conference on Neural Networks 2006 (IJCNN'06)*, age 4139 4145. IEEE, 200 .

21 ei . a e ian ing ang mac ine, cl e ing and n mbe f cl e . *Pattern Recognition Letters*, 18 11 11 7 1178, 1997.

22 i and in en Ma. A g adien en eg la- i ed li eli d lea ning alg i m n ga ian mi e i a ma ic m del elec i n. In *Advances in Neural Networks-ISNN 2006*, age 4 4 4 9. S inge , 200 .

23 i and ace -S ing I . en e ali ed c m- e i i e lea ning f ga ian mi e m del . *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39 4 901 909, 2009.

24 ei , Adam a , and E i O a. i al e- n ali ed c m e i i e lea ning f cl e ing anal i , bf n e , and c e de ec i n. *Neural Networks, IEEE Transactions on*, 4 4 3 49, 1993.

25 in en Ma and ai n ang. A c -f nc i n a - ac i al en ali ed c m e i i e lea ning cl . *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 3 4 722 737, 200 .

2 id ng in, Man- e ng ng, and . S. e ng. Scalable m del-ba ed cl e ing f la ge da aba - e ba ed n da a mma i a i n. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27 11 1710 1719, 2005.

27 in en Ma and ian feng i . e b annealing lea n- ing alg i m f ga ian mi e i a ma ed m del elec i n. *Pattern Recognition*, 40 7 2029 2037, 2007.

28 UCI Mac ine ea ning e i . www .ic . ci.ed /~mlea n/.

29 Ian lliffe. *Principal Component Analysis*. ile Online ib a , 2005.