DYNAMICALLY REGULARIZED MAXIMUM LIKELIHOOD LEARNING OF GAUSSIAN MIXTURES

Jinwen Ma, Hongyan Wang

Department of Information Science, School of Mathematical Sciences
Peking University, Beijing, 100871, China

ABSTRACT

The Gaussian mixture model is widely applied in the fields of data analysis and information processing. Recently, its parameter learning with adaptive model selection, i.e., the adaptive selection of number of Gaussian distributions in the mixture for a given sample dataset, has become an attracting and interesting topic. In this paper, we propose a dynamically regularized maximum likelihood learning (DRMLL) algorithm for Gaussian mixtures with adaptive model selection. The basic idea is that the Bayesian Ying-Yang (BYY) harmony learning is interpreted as the maximum likelihood learning regularized by the average Shannon entropy of the posterior probability per sample scaled by a positive parameter. As this scale parameter dynamically decreases from 1 to 0, the DRMLL algorithm transforms from the BYY harmony learning with adaptive model selection to the final maximum likelihood (ML) learning. It is demonstrated by simulation experiments that the DRMLL algorithm can not only select the correct number of actual Gaussian distributions in a dataset, but also obtain ML estimates of the parameters in the original mixture.

Keywords: Gaussian mixtures; BYY Harmony learning; Adaptive model selection; Regularization; Maximum likelihood.

1. INTRODUCTION

As a powerful tool for data analysis and information processing, the Gaussian mixture model has been widely applied to data modeling and clustering analysis on a given dataset. Actually, there have been various statistical learning methods to estimate the parameters in the mixture, e.g., the EM algorithm [1] and the method of moments [2]. Usually, it is assumed that the number of Gaussians or clusters in the dataset is pre-known. However, in many instances this key information is not available and thus the selection of an appropriate number of Gaussians must be made before or during the estimation of the parameters in the mixture, which is a rather complicated and difficult task [3].

Jinwen Ma, the corresponding author, Telephone:86-10-62760609, E-mail: jwma@math.pku.edu.cn.

As the number k of Gaussians is just a scale of the Gaussian mixture model, its determination is often referred to as model selection. In this light, the general Gaussian mixture modeling is essentially a compound problem of parameter estimation and model selection. Actually, this compound problem has been investigated in different ways. The traditional method is to choose an optimal number of Gaussians via certain selection criterion. Among these criteria, Akaike's Information Criterion (AIC) [4] and Bayesian Inference Criterion (BIC) [5] are well-known. But the validating process is computationally consumptive because we need to repeat the entire parameter learning process at a large number of possible values of k.

Since the 1990s, some new statistical learning approaches have been proposed to solve this compound problem. The Dirichlet processes [6] and reversible jump Markov chain Monte Carlo (RJMCMC) [7] are two typical stochastic simulation methods for inferring the optimal mixture, which generally require a large number of samples through different sampling rules. The variational Bayesian approach [8] tries to make a Bayesian model search via optimizing a variational bound, which is still lack of rigorous theoretical support. The unsupervised learning approach [9] introduces some competitive learning mechanism into the EM algorithm such that model selection can be made adaptively during the parameter learning by using a simplified MML criterion.

From a new view of Bayesian statistical learning, the Bayesian Ying-Yang (BYY) harmony learning [10]-[11] has also provided a useful approach to making model selection adaptively during parameter learning. In fact, it has already been implemented on Gaussian mixtures as well as the other finite mixture and several BYY harmony learning algorithms (e.g., [12]-[15]) have already been established and demonstrated well on adaptive model selection. Although the BYY harmony learning is capable of performing adaptive model selection, its parameter estimates have certain deviation from the ML parameter estimates which are consistent with the true parameters. So, it is better to combine the functions of adaptive model selection and the consistent estimates of the parameters together into one unified learning framework.

978-1-4799-2186-7/14/$31.00 ©2014 IEEE

Based on the analysis of BYY harmony function on Gaussian mixtures under the BI-architecture [14]-[15], the BYY harmony learning can be regarded as a kind of regularization version of the maximum likelihood (ML) learning. The regularization term is just the average Shannon entropy of the posterior probabilities per sample. In fact, the entropy regularization methods could date back to the 1980s and have been widely used in the ill-posed problems involving in model selection. In the view of model selection and ML parameter estimation, the BYY harmony function can be decomposed into the likelihood function and the entropy regularized term. However, if the regularization scale keeps constant in the way of the existing regularized ML learning approaches [16]-[18], the regularized ML learning leads to a certain deviation between its estimation and the ML or true solution. In order to overcome this problem, we can adjust the regularization scale from 1 to 0, the learning process can transform from the BYY harmony learning into the ML learning. If we further maintain the increase of the regularization scale dynamically and properly, the regularized learning process can lead to the ML estimates of the parameters with adaptive model selection on Gaussian mixtures. Oppositely, the ML learning can be regarded as a kind of regularization version of the BYY harmony learning and the dynamically regularized BYY harmony learning algorithm can be established for Gaussian mixtures [20].

In the current paper, we propose a Dynamically Regularized Maximum Likelihood Learning (DRMLL) algorithm for Gaussian mixtures with adaptive model selection. By controlling the regularization scale to dynamically decrease from 1 to 0, the DRMLL algorithm transforms from the BYY harmony learning with adaptive model selection to the conventional maximum likelihood learning. It is demonstrated by the experiments that the DRMLL algorithm can not only select the correct number of actual Gaussian distributions in a given dataset, but also obtain ML estimates of the parameters in the original mixture.

2. DRMLL ALGORITHM

In this section, we firstly present the dynamic regularization mechanism to be used. Then, we introduce the fixed-point algorithm for the dynamic learning process. We further discuss the dynamic evolution of the regularization scale factor. Finally, we give the complete DRMLL algorithm.

2.1. Dynamic Regularization Mechanism

According to [15], for the Gaussian mixture model $P(x|\Theta_k) = \sum_{j=1}^{k} \pi_j q(x_t|m_j, \Sigma_j)$, the corresponding BYY harmony function $J(\Theta_k)$ can be divided into two parts,

$$J(\Theta_k) = L(\Theta_k) - O_N(p(y|x)), \tag{1}$$

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

$$L(\Theta_k) = \frac{1}{N}\sum_{t=1}^{N} \ln(\sum_{j=1}^{k}(\pi_j q(x_t|m_j, \Sigma_j))), \tag{2}$$

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

$$O_N(p(y|x)) = -\frac{1}{N}\sum_{t=1}^{N}\sum_{j=1}^{k} p(j|x_t)\ln p(j|x_t). \tag{3}$$

According to Eq.(1), if $-O_N(p(y|x))$ is viewed as a regularization term, the BYY harmony learning, i.e., maximizing $J(\Theta_k)$, is a regularized ML learning which has already been investigated in [17, 18] by scaling the regularization term with a small positive number. However, since they keep the regularization scale constant just as in the case of the BYY harmony learning, these approaches must suffer from inconsistent parameter estimation.

To dynamically control the regularization, we use a dynamic regularization scale factor $\lambda(\geq 0)$ and have

$$J_\lambda(\Theta_k) = L(\Theta_k) - \lambda O_N(p(y|x)). \tag{4}$$

If $\lambda = 1$, $J_\lambda(\Theta_k) = J(\Theta_k)$ is just BYY harmony function on the Bi-architecture for Gaussian mixtures. If $\lambda = 0$, $L_\lambda(\Theta_k)$ is the log-likelihood function of the Gaussian mixture model. That is, with λ decreasing from 1 to 0, maximizing $J_\lambda(\Theta_k)$ changes from the BYY harmony learning to the ML learning. Here we try to control the decreasing of λ dynamically and appropriately to realize adaptive model selection at the previous learning stage and the ML estimation at the final learning stage.

2.2. Fixed-point Learning Algorithm

At each phase of the dynamically regularized maximum likelihood learning with a particular λ, we construct a fixed-point algorithm to maximize $J_\lambda(\Theta_k)$ as follows.

For convenience, we utilize the softmax representation for π_j, i.e., $\pi_j = e^{\beta_j}/\sum_{i=1}^{K} e^{\beta_i}$, $j = 1, \cdots, k$, where $\beta_j \in (-\infty, +\infty), j = 1, \cdots, k$. Letting the derivatives of $J_\lambda(\Theta_k)$ with respect to β_j, m_j and Σ_j, respectively, be zero, we get the following fixed-point (iterative) learning algorithm:

$$\hat{\pi}_j = \frac{\sum_{t=1}^{N} p(j|x_t)\gamma_j(t)}{N}; \tag{5}$$

$$\hat{m}_j = \frac{\sum_{t=1}^{N} p(j|x_t)\gamma_j(t)x_t}{\sum_{t=1}^{N} p(j|x_t)\gamma_j(t)}; \tag{6}$$

$$\hat{\Sigma}_j = \frac{\sum_{t=1}^{N} p(j|x_t)\gamma_j(t)(x_t - \hat{m}_j)(x_t - \hat{m}_j)^T}{\sum_{t=1}^{N} p(j|x_t)\gamma_j(t)}, \tag{7}$$

where

$$\gamma_j(t) = 1 + \lambda \ln p(j|x_t) - \lambda \sum_{i=1}^{k} p(i|x_t)\ln p(i|x_t). \tag{8}$$

In comparison with the conventional EM algorithm [1], this fixed-point learning algorithm differs only at the augmenting term $\gamma_j(t)$. It can be easily verified that when $\lambda = 0$, $\gamma_j(t) = 1$, the fixed-point learning algorithm is just the EM algorithm and when $\lambda = 1$, the fixed-point learning algorithm returns to the original fixed-point BYY learning algorithm [15] for maximizing the harmony function $J(\Theta_k)$.

Actually, $\gamma_j(t)$ implements a rival penalized competitive learning (RPCL) mechanism [19] so that model selection can be made adaptively during parameter learning. At the early learning stage, $\gamma_j(t) < 0$ may happen. According to Eq.(8), the mean vectors of the j-th Gaussian will move away from x_t. Otherwise, if $\gamma_j(t) > 0$, the mean vectors of the j-th Gaussian will be attracted to x_t. So, for x_t, Gaussians with $\gamma_j(t) > 0$ are winners while these Gausians with $\gamma_j(t) < 0$ are losers.

However, the fixed-point learning algorithm cannot guarantee the positive definiteness of each covariance matrix during the iteration since $\gamma_j(t)$ may be negative. In order to overcome this problem, we use the EM update rule of the covariance matrixes, i.e., forcing all $\gamma_j(t) = 1$ in Eq.(7), in this degenerated case. In fact, this simplification is applicable and efficient since the competition for adaptive model selection is mainly among mean vectors and controlled by the mixing proportions.

2.3. Dynamic Evolution of λ

We further discuss the dynamic evolution of λ with time T during the learning process. According to our regularization mechanism, λ should start around 1 and decrease slowly at the early learning stage to realize adaptive model selection. Then, at the sequent stage, λ can attenuate to 0 at a higher speed so that the algorithm will finally converge to a ML solution. So, it is crucial to check whether the adaptive model selection has accomplished and when to change learning stage.

In order to detect the turning point, we introduce the Shannon entropy of mixing proportions in the Gaussian mixture model, $H_\pi = -\sum_{j=1}^{k} \pi_j \ln \pi_j$. It is obvious that H_π is sensitive to the structure of the Gaussian mixture model. If model selection is not completed, the difference of H_π between two iterations is considerable. Otherwise, the difference should be very small. This motivates us to adopt the absolute change rate of H_π between two iterations, defined by

$$h_\pi(T) = |\frac{H_\pi(T) - H_\pi(T-1)}{H_\pi(T)}|, \tag{9}$$

as an indicator of model selection. Here, T is the time, i.e., the number of iterations. The whole learning process is divided into two learning stages according to a given threshold $\varepsilon_1(> 0)$ of this indicator. That is, if $h_\pi(T) > \varepsilon_1$, $\lambda(T)$ increases at a low speed; otherwise, it increases at a high speed. Since $\lambda(T)$ is assumed to increase exponentially, its dynamic evolution process can be given as follow:

$$\lambda(T) = \begin{cases} 1 - \lambda_0 * \eta_1^T, & \text{if } h_\pi(T) > \varepsilon_1; \\ 1 - \lambda_0 * (\frac{\eta_1}{\eta_2})^{T^*} \eta_2^T, & \text{if } h_\pi(T) \le \varepsilon_1, \end{cases} \tag{10}$$

where λ_0 is a very small positive constant, η_1, η_2 are two positive constants with the constraint that $1 < \eta_1 < \eta_2$, and T^* is the turning point such that $h_\pi(T^*) > h_0$ and $h_\pi(T^* + 1) \le h_0$. When λ becomes 0, we fix it until the algorithm stops.

2.4. Complete DRMLL Algorithm

We finally summary our proposed DRMLL algorithm. Firstly, we should choose the parameters of the algorithm properly. As mentioned previously, λ_0, η_1, η_2 and ε_1 must be carefully selected to make the evolution of $\lambda(T)$ dynamic. θ_0 is a threshold value to filter out Gaussians with very small mixing proportions during the parameter learning process, while $\varepsilon_2(> 0)$ is a threshold value to terminate the iteration. If $\lambda = 0$ and the absolute increment of the log likelihood is smaller than ε_2, we affirm the convergence of the algorithm. In our learning paradigm, k is flexible. However, it should be larger than the number k^* of actual Gaussians or clusters in the dataset. As for the initial setting of the parameters Θ_k, i.e., $\Theta_k^{(0)} = \{\pi_i^0, m_i^0, \Sigma_i^0\}|_{i=1}^k$, some competitive learning mechanism may be helpful. For example, m_i^0 can be selected through a DSRPCL procedure [19] and then π_i^0 and Σ_i^0 can be estimated accordingly.

After initializing all the parameters, Θ_k will be updated in each phase of $\lambda(T)$ via the fixed-point learning algorithm given by Eqs (12)-(14). At the end of each learning phase, the Gaussians with the mixing proportions less than θ_0 are annihilated immediately. After $\lambda(T)$ becomes 0, the algorithm goes on until the log likelihood function reaches its maximum value or its absolute increment is less than ε_2.

3. EXPERIMENTAL RESULTS

In this section, various experiments are carried out to demonstrate the performance of the DRMLL algorithm for Gaussian mixtures. Moreover, it is compared with some typical existing learning algorithms. In these experiments, we always select $\varepsilon_1 = 1e-5$, $\varepsilon_2 = 1e-5$, $\eta_1 = 1.005$,$\lambda_0 = 1e-5$ $\eta_2 = 2$ and $\theta_0 = 0.05$. The other parameters will be specified in the particular experiments.

We begin to generate four typical synthetic datatsets from mixtures of four or three bivariate Gaussian distributions on the plane coordinate system (i.e., $d = 2$). Clearly, these

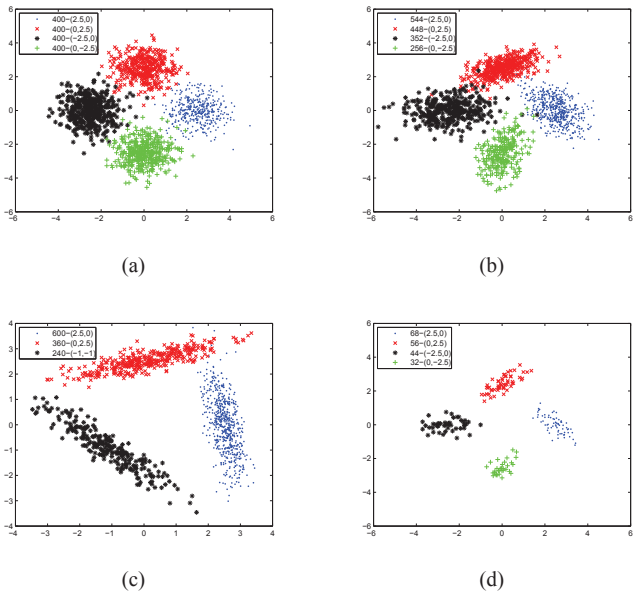

Fig. 1. Four synthetic datasets for simulation experiments. (a). $\mathcal{S}_1$, (b). $\mathcal{S}_2$, (c). $\mathcal{S}_3$, (d). $\mathcal{S}_4$.

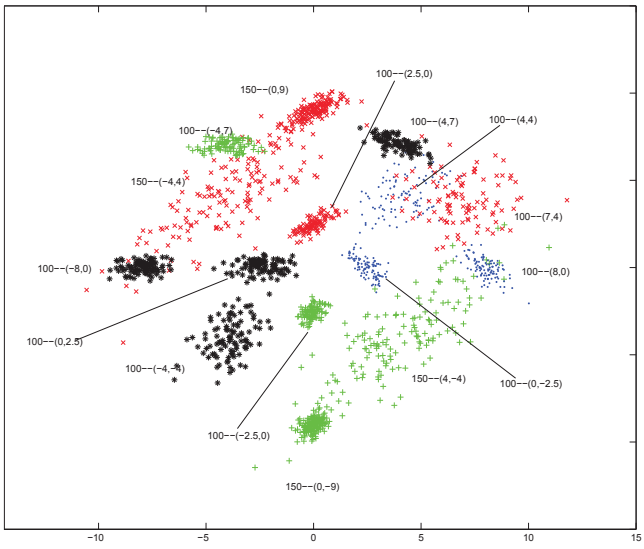

Fig. 2. Synthetic dataset $\mathcal{S}_5$ of 15 Gaussians with different elliptical shapes and unequal mixing proportions.

Gaussian distributions are either sphere-shaped or ellipse-shaped. As shown in Fig.1, the covariance matrices of Gaussian distributions are designed to demonstrate different degrees of overlap among Gaussians (i.e., clusters). Moreover, the four datasets are also generated with equal or unequal mixing proportions. The specific parameters for these four datasets are listed in Table 1, where m_i, $\Sigma_i = (\sigma_{jk}^i)_{2\times 2}$, π_i, N_i denote the mean vector, covariance matrix, mixing proportion, and number of samples of the i-th Gaussian, respectively.

The DRMLL algorithm is implemented on each of these four synthetic datasets with $k = 2k^*$. Moreover, we compare the DRMLL algorithm with the MML-based unsupervised learning algorithm particularly for Gaussian mixtures [9], being referred to as CEM^2 for short. Actually, CEM^2 has been considered as a typical and competitive learning algorithm for the Gaussian mixture learning with adaptive model selection in literature. To show the stability and accuracy of converged results, we implement both CEM^2(with the stop criterion $\epsilon = 10^{-6}$) and the DRMLL algorithm on each of the four datasets for 50 times with different randomly selected initial parameters. We then compute the frequencies of correct model selection (CMS) and average runtime of these two algorithms over 50 trials on each dataset. The experimental results are listed in Table 2. Obviously, the DRMLL algorithm considerably outperforms CEM^2 on both correct model selection and runtime.

In addition to model selection and runtime, we also compare the DRMLL algorithm with CEM^2 on the accuracy of parameter estimation. For each parameter θ_i, we define $\Delta\theta_i$ as the average absolute error of $|\theta_i - \theta_i^*|$ over 50 trials. For each dataset, we compute the total average absolute error per each parameter called TAE. Actually, the TAEs of the two algorithms on the four datasets are listed in Table 3. It can be found that the DRMLL algorithm and CEM^2 have almost the same accuracy on parameter estimation. However, for the fourth dataset with a small number of samples, the accuracy of the DRMLL algorithm is remarkably better than that of CEM^2.

Table 3. The comparison of the DRMLL algorithm with CEM^2 on parameter estimation accuracy.

Dataset	DRMLL	CEM^2
$\mathcal{S}_1$	0.0204	0.0204
$\mathcal{S}_2$	0.0171	0.0172
$\mathcal{S}_3$	0.0363	0.0363
$\mathcal{S}_4$	0.0251	0.0715

The DRMLL algorithm is also compared with the BYY annealing algorithm (BYY-AEM) [14]. While the DRMLL and BYY-AEM algorithms have the similar performance on adaptive model selection, the DRMLL algorithm leads to a more accurate parameter estimation. Actually, the TAEs of the BYY-AEM algorithm on the four datasets are respectively 0.0204, 0.0243, 0.0386 and 0.0322, which are mostly higher than those of the DRMLL algorithm.

Finally, we implement the DRMLL algorithm on a rel-

Table 1. The values of the parameters of the four synthetic datasets.

The dataset	Gaussian	m_i	σ_{11}^i	$\sigma_{12}^i(\sigma_{21}^i)$	σ_{22}^i	π_i	N_i
$\mathcal{S}_1$	G1	(2.50,0)	0.50	0.00	0.50	0.25	400
(N=1600)	G2	(0,2.50)	0.50	0.00	0.50	0.25	400
	G3	(-2.50,0)	0.50	0.00	0.50	0.25	400
	G4	(0,-2.50)	0.50	0.00	0.50	0.25	400
$\mathcal{S}_2$	G1	(2.50,0)	0.45	-0.25	0.55	0.34	544
(N=1600)	G2	(0,2.50)	0.65	0.20	0.25	0.28	448
	G3	(-2.50,0)	1.00	0.10	0.35	0.22	352
	G4	(0,-2.50)	0.30	0.15	0.80	0.16	265
$\mathcal{S}_3$	G1	(2.50,0)	0.10	-0.20	1.25	0.50	600
(N=1200)	G2	(0,2.50)	1.25	0.35	0.15	0.30	360
	G3	(-1,-1)	1.00	-0.80	0.75	0.20	240
$\mathcal{S}_4$	G1	(2.50,0)	0.28	-0.20	0.32	0.34	68
(N=200)	G2	(0,2.50)	0.34	0.20	0.22	0.28	56
	G3	(-2.50,0)	0.50	0.04	0.12	0.22	44
	G4	(0,-2.50)	0.10	0.05	0.50	0.16	32

Table 2. The comparison of the DRMLL and CEM^2 algorithms on model selection and runtime.

Datasets	DRMLL		CEM^2	
	CMS Frequency	runtime(s)	CMS Frequency	runtime(s)
$\mathcal{S}_1$	100%	707	84%	11290
$\mathcal{S}_2$	100%	764	56%	1825
$\mathcal{S}_3$	100%	405	72%	4317
$\mathcal{S}_4$	98%	250	56%	554

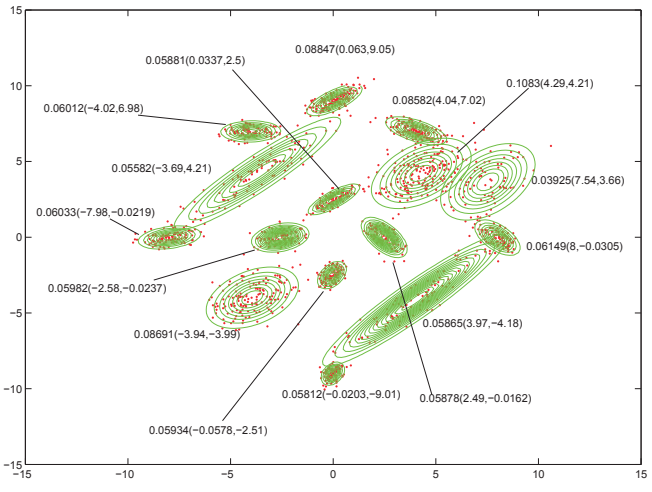

Fig. 3. The adaptive model selection result of the DRMLL algorithm on $\mathcal{S}_5$.

atively complicated dataset $\mathcal{S}_5$ of 15 Gaussians with different elliptical shapes and unequal mixing proportions, being sketched in Fig. 2. It can be observed from Fig. 3 that the DRMLL algorithm can still make correct model selection adaptively in this complicated case, even if there are some actual Gaussians which are very flat and strongly overlapped. Actually, this complicated structure makes model selection be very difficult. Therefore, the experimental result on this dataset further demonstrates that the DRMLL algorithm owns a good ability of adaptive model selection. Moreover, as the DRMLL process tends to be the maximum likelihood learning at the final stage, the parameter estimation is as good as the ML solution.

4. CONCLUSIONS

We have investigated the relationship between the BYY harmony learning and the ML learning and found out that the BYY harmony learning can be regarded as the maximum likelihood learning being regularized by the average Shannon entropy of the posterior probability per sample. Using the dynamic regularization mechanism, the dynamically regularized maximum likelihood learning (DRMLL) algo-

rithm is established for Gaussian mixtures. By controlling the scale factor of the regularization term to dynamically increase from 1 to 0, the DRMLL algorithm starts from the BYY harmony learning with a capability of adaptive model selection, and then gradually transforms to the conventional maximum likelihood learning to obtain a consistent parameter estimation. It is demonstrated by simulation experiments that the DRMLL algorithm can not only select the correct number of actual Gaussians in a dataset, but also obtain the ML estimates of the parameters in the mixture.

Acknowledgments.

This work was supported by the Natural Science Foundation of China for Grants 61171138 and 60771061.

5. REFERENCES

[1] R. A. Render, H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol. 26, no.2, pp. 195-293, 1984.

[2] N.E. Day, "Estimating the components of a mixture of normal distributions," *Biometrika*, vol.56, pp. 463-474, 1969.

[3] J. A. Hartigan, "Distribution problems in clustering," *Classification and Clustering*, J. Van Ryzin Eds., pp. 45-72, New York: Academic Press, 1977.

[4] H. Akaike. "A New Look at the Statistical Model Identification," *IEEE Trans. on Automatic Control*, vol. AC-19, pp. 716-723, 1974.

[5] G. Scharz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, pp. 461-464, 1978.

[6] M.D. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association*, vol.90, no.430, pp. 577-588, 1995.

[7] S. Rechardson and P. J. Green, "On Bayesian analysis of mixtures with an unknown number of components," *Journal of the Royal Statistical Soceity B*, vol.59, no.4, pp. 731-792, 1997.

[8] N. Ueda and Z. Ghahramani, "Bayesian model search for mixture models based on optimizing variational bounds," *Neural Networks*, vo.15, no.10, pp. 1223-1241, 2002.

[9] M.A.F. Figueiredo and A.K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.24, no.3, pp. 381-396, 2002.

[10] L. Xu, "Best harmony, unified RPCL and automated model selection for unsupervised and supervised learning on Gaussian mixtures, three-layer nets and ME-RBF-SVM models," *International Journal of Neural Systems*, vol.11, no.1, pp. 43-69, 2001.

[11] L. Xu, "BYY harmony learning, structural RPCL, and topological self-organizing on mixture models," *Neural Networks*, vol.15, pp. 1125-1151, 2002.

[12] J. Ma and T. Wang, L. Xu, "A gradient BYY harmony learning rule on Gaussian mixture with automated model selection," *Neurocomputing*, vo.56, pp. 481-487, 2004.

[13] J. Ma and L. Wang, "BYY harmony learning on finite mixture: adaptive gradient implementation and a floating RPCL mechanism," *Neural Processing Letters*, vol.24, no.1, pp. 19-40, 2006.

[14] J. Ma and J. Liu, "The BYY annealing learning algorithm for Gaussian mixture with automated model selection," *Pattern Recognition*, vol.40, pp. 2029-2037, 2007.

[15] J. Ma and X. He, "A fast fixed-point BYY harmony learning algorithm on Gaussian mixture with automated model selection," *Pattern Recognition Letters*, vol.29, no.6, pp. 701-711, 2008.

[16] J. Ma and T. Wang, "Entropy penalized automated model selection on Gaussian mixture," *International Journal of Pattern Recognition and Artificial Intelligence*, vo.18, no.8, pp. 1501-1512, 2004.

[17] Z. Lu and J. Ma, "A gradient entropy regularized likelihhood learning algorithm on Gaussian mixture with automatic model selection," *Lecture Notes on Computer Science*, vol.3971, pp. 464-469, 2006.

[18] Z. Lu and H.S. Ip, "General competitive learning of Gaussian mixture models," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol.39, no.4, pp. 901-909, 2009.

[19] J. Ma and T. Wang, "A cost-function approach to rival penalized Competitive learning (RPCL)," *IEEE Trans. on Systems, Man and Cybernetics, Part B: Cybernetics*, vol.36, no.4, pp. 722-737, 2006.

[20] H. Wang and J. Ma, "Dynamically regularized harmony learning of Gaussian mixtures," Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC2014), Oct. 5-8, 2014, San Diego, CA, USA.