

International Journal of Pattern Recognition and Artificial Intelligence
Vol. 18, No. 8 (2004) 1501–1512
(c) World Scientific Publishing Company

# ENTROPY PENALIZED AUTOMATED MODEL SELECTION ON GAUSSIAN MIXTURE\*

#### JINWEN MA

Department of Information Science, School of Mathematical Sciences & LMAM, Peking University, Beijing, 100871, P. R. China jwmamath.pku.edu.cn

#### TAIJUN WANG

Department of Wireless Engineering, Southeast University, Nanjing, 210018, P. R. China

Gaussian mixture modeling is a powerful approach for data analysis and the determination of the number of Gaussians, or clusters, is actually the problem of Gaussian mixture model selection which has been investigated from several respects. This paper proposes a new kind of automated model selection algorithm for Gaussian mixture modeling via an entropy penalized maximum-likelihood estimation. It is demonstrated by the experiments that the proposed algorithm can make model selection automatically during the parameter estimation, with the mixing proportions of the extra Gaussians attenuating to zero. As compared with the BYY automated model selection algorithms, it converges more stably and accurately as the number of samples becomes large.

*Keywords*: Gaussian mixture; model selection; maximum-likelihood estimation; Shannon entropy; penalty.

#### 1. Introduction

esta antesta de t	· ·· <u>-</u> · · · · · ·	· · · <u>-</u> · <u>-</u> ·	· · · · - ·	
	··· / · · <u>·</u>	. K. 4. (1997) - 1997	· -,· · · · - · · , -	- • • • • • •
	<u> </u>			
	$\sim - \cdots + - \cdots$	- / /		
··· · / ··· <u>_</u> ·	·· · · · · · -	17	K	
· · · · · · · · · · · · · · · · · · ·	2	· · · · · · ·		· · · _ · _ · _ · _ · _ · _ · _ · _ · _
· _ · · · · ·	?.	9		<i></i>
··· · · · · - · ·	· _ · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · - · ·	- / ./
· · · · · · · · · · · ·	/ . / <b>/</b>		· · · · · · · ·	
	· · · · · · · · · · · · · · · · · · ·			- · _ · ·

\*This work was supported by the Natural Science Foundation of China for Project 60071004.

K 、
8
$\begin{array}{cccccccccccccccccccccccccccccccccccc$
V*9
$\dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots $
···
$\cdots \cdots $
· _ · · · · · · · · · · · · · · · · · ·
$\cdots \qquad \qquad$
Κ
···· · · · · · · · · · · · · · · · · ·
the second s
22
en en antal en en en antal de la companya de la com
· · _/
·····K 、 ······························
24
$\cdot  \cdot  \cdot  \cdot  \cdot  \cdot  \cdot  \cdot  \cdot  \cdot $
23 27 27 27 27 27 27 27 27 27 27 27 27 27
· · · · · · · · · · · · · · · · · · ·
···· <u>~ -</u> ··· ··· · <u>-</u> ··· · · · · · · · · · · · · · · · · ·
· <u>·</u> ··································
· _ · · · _ · _ · _ · · · · · · · · · ·
· · · · · · · · · · · · · · · · · · ·
· · · · · · · · · · · · · · · · · · ·
······································
· · · · · · · · · · · · · · · · · · ·

, · · · , · · · · · ... · · <u>--</u> · · . . \_/ . . . . · · \_  $\cdots = \cdots = \cdots = \cdots = \cdots = \cdots = \cdots = \cdots$ 

Entropy Penalized MLE Iterative Algorithm

 $\mathbf{P} \mathbf{x} \mathbf{j} = \sum_{i=1}^{K} j \mathbf{P} \mathbf{x} \mathbf{j} \mathbf{m}_{j} \mathbf{\hat{x}} j; \quad j = i; \quad \sum_{i=1}^{K} j = i$ 2  $2 \frac{1}{2} e^{-\frac{1}{2}(x-m_j)^T \sum_{j=1}^{j-1} (x-m_j)} e^{-\frac{1}{2}(x-m_j)^T \sum_{j=1}^{j-1} (x-m_j)}$  $\mathbf{P}$  xjm  $\hat{\mathbf{S}}_{i}$ ?  $\dots \quad \underline{K} \quad \underline{\lambda} \quad \underline{\lambda$ **X:**  $m_i$   $m_i$  j · ·\_. · · · · · · · · · *j* . . . ·· \_  $\sum_{j=1}^{K} j$ . . **j** . . j· · · · · <u>-</u> · <u>-</u> · · · · **.** . . . . . . . . <u>. . .</u> . · \_ · \_ · P \_\_\_\_\_. . \_ . · · · · · <u>--</u> ·· \_ · · · · · · <u>-</u> · · · · · · \_ 2 ·· - · - · · · · · · · · · · · · <u>-</u> · · · · · · · · · · · · · . . . .,  $\mathbf{S} = \mathbf{f} \mathbf{x}_t \mathbf{g}_{t=1}^N$ P∕xi

 $\mathbb{D} \qquad \sum_{t=1}^{N} \quad \mathbb{P} \mathbf{x}_{t} \mathbf{j} : \qquad 2$ 



@0 Td ( 6.(es)Tj 8701Tj /R32 9.96264 Tf 6.83435 0 Td (( )Tj /R32 9.96264 Tf 11.6264 0 Td (;)Tj 4.43962 0 Td ( )Tj /



. . .

?

?

$$\frac{\mathbf{N}}{\mathbf{N}} = 2 \mathbf{n} \mathbf{P} \left( \sum_{t=1}^{N} \mathbf{h}_{j}^{2} \mathbf{t} \mathbf{N} \mathbf{N} \mathbf{h}_{j}^{2} \mathbf{h}_{j}^{$$

$$\mathbf{m}_{j}^{+} = \frac{1}{\sum_{t=1}^{N} \mathbf{h}_{j}^{2} \mathbf{t}} \sum_{t=1}^{N} \mathbf{h}_{j}^{2} \mathbf{t} \mathbf{x}_{t}$$
 2

. . . . . .

$$\mathbf{\hat{\mathbf{x}}}_{j}^{+} = \frac{1}{\sum_{t=1}^{N} \mathbf{h}_{j}^{2} \mathbf{t}} \sum_{t=1}^{N} \mathbf{h}_{j}^{2} \mathbf{t}^{2} \mathbf{x}_{t} = \mathbf{m}_{j}^{2} \mathbf{x}_{t} = \mathbf{m}_{j}^{-T} \mathbf{t}^{T} \mathbf{t}^{2} \mathbf{x}_{t}$$

### 3. Simulation Results

· · .			-	· · · = · · =			· · <u>-</u> · · ·	· · — ·
· · · · ·		· · · · · ·		· _ · ·			· / · · <del>- ·</del>	· _ · ·
- <b>-</b>	· ·		· · · 、—	· · · ·	- / / /	· _		· · - · · ·
· · _/		· · _ · · ·	,			· · _/		

## 3.1. The sample data





### 2. Automated model selection





1510 J. Ma & T. Wang

## **3.3.** Parameter estimation

#### eferences

- S. C. Ahalt, A. K. Krishnamurty, P. Chen and D. E. Melton, Competitive learning algorithm for vector quantization,  $e_{4}$ ,  $e_{7}$ ,  $e_{7}$ ,  $a_{3}$ (3) (1990) 277–291. H. Akaike, A new look at the statistical model identification,  $\tilde{a}_{77}$ ,  $\tilde{n}$ ,  $A_{4}$  on on  $\tilde{n}$ , **AC-19** (1974) 716–723.
- J. C. Bezdek and N. R. Pal, Some nex indexes of cluster validityContr.

- 21. X. L. Xie and G. A. Beni, A validity method for fuzzy clustering,  $\tilde{n} = n$ An,  $M c = n \epsilon$ , **13**(8) (1991) 843-847.
- 22. L. Xu, A. Krzyżak and E. Oja, Rival penalized competitive learning for clustering analysis, RBF net, and curve detection, n = n = n = 1 (1993) 636-648.
- 23. L. Xu e A unified learning scheme: Bayesian-Kullback YING-YANG machine, in Ad nee n extra provide on roce n extra provide no roce n provide no roce n provide no roce n, pp. 977–988.
- 24. L. Xu, Rival penalized competitive learning, finite mixture, and multisets clustering, in roc 1 in on one one of one of the or , Vol. 3, pp. 251-253.
- 25. 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,  $n = e_{4} + e_{7} + 11(1) (2002) 43-69$ .
- L. Xu, BYY harmony learning, structural RPCL, and topological self-organzing on mixture modes, e<sub>4</sub>, e o<sup>-</sup> 15 (2002) 1231–1237.
- 27. L. Xu, Ying-Yang learning, in  $\bullet$  nd oo of  $B^n$  n  $\bullet o^n y$  nd  $\bullet_{q'}$   $\bullet$  on , ed. M. A. Arbib, 2nd edn. (The MIT Press, 2002), pp. 1231–1237.



Jinwen Ma received the M.S. degree in applied mathematics from Xi'an Jiaotong University in 1988 and the Ph.D. in probability theory and statistics from Nankai University in 1992. From July 1992 to November 1999, he

was a Lecturer and Associate Professor at the Department of Mathematics, Shantou University. He has been a full Professor at the Institute of Mathematics, Shantou University since December, 1999. In September 2001, he was transferred to the Department of Information Science at the School of Mathematical Sciences, Peking University. During 1995 and 2003, he also made several visits to the Department of Computer Science and Engineering, the Chinese University of Hong Kong as a Research Associate or Fellow.

He has published over 40 academic papers on neural networks, pattern recognition, artificial intelligence and information theory.



Taijun Wang received the M.S. degree in signal and information processing from Southeast University, China, in 1982. He has been an Associate Professor at the Department of Radio Engineering of Southeast University

since 1988.

His research interests are related to digital signal processing, pattern recognition, artificial neural networks, statistical learning theory and scientific visualization.