

1. Introduction

2. The EM algorithm for mixtures of densities from exponential families

2.1. The mixture model

$$
P(x|F) = \sum_{i=1}^{K} a_i P_i(x|f_i), \quad a_i \ge 0, \sum_{i=1}^{K} a_i = 1,
$$
\n
$$
x = [x_1, ..., x_n] \in R^n, \quad P_i
$$
\n
$$
r = \begin{cases} r_i, & r_i \in O_i \subset R^{di}, & R_i \in O_i \subset R^{di}, \\ r_i, & r_i \in (a_1, ..., a_K, f_1, ..., f_K) \in O, \\ r_i, & r_i \in (a_1, ..., a_K, f_1, ..., f_K) \in O, \\ r_i(x|f_i) = P_i(x|m_i, S_i) & G_i \end{cases}
$$
\n
$$
P_i(x|f_i) = P_i(x|m_i, S_i) = \frac{1}{(2p)^{n/2}(-S_i)^{1/2}} \quad \text{(1/2)(x-mi) } S_i^{-1}(x-mi),
$$
\n
$$
m_i = [m_{i1}, ..., m_{in}] \quad S_i = (s_{ki}^i)_{n \times n}
$$
\n
$$
S_i = (s_{ki
$$

 $\begin{array}{ccc} 0 & , & \end{array}$

$$
P(x|f) = q(x|y(f)) = a(f)^{-1}b(x) {^{y(f) t(x)}}, \quad x \in R^n
$$

\n
$$
\vdots
$$

\n
$$
\vdots
$$

\n
$$
\vdots
$$

\n
$$
\vdots
$$

\n
$$
I = E_y(t(X))
$$

$$
P(x|f) \t m \t U(x|f) \t S,
$$

\n
$$
P(x|f) \le U(x|f) = w(x)(1) e^{-c_1 - r(1/\mathbb{Z}(x))^{-c_2}},
$$
\n(4)

^ZðxÞ ¼ ^ðlmax^Þ n kx mk and lmax is the maximum eigenvalue of S of PðxjfÞ: Moreover, c1; c2; r and n are a group of positive numbers, and wðxÞ is a positive polynomial function of x1; ... ; xn with constant coefficients. Here and hereafter, we use the Euclidean norm for a vector and its inductive norm for a matrix. Actually, the class of these families

G
\n
$$
\begin{array}{ccc}\n\mathbf{G} & & P_i(x|\mathbf{f}_i) \\
\mathbf{f}_i \in \mathbf{O}_i \subset R^{d_i} & \vdots \\
P_i(x|\mathbf{f}_i) = a_i(\mathbf{f}_i)^{-1}b_i(x) & \mathbf{y}_i(\mathbf{f}_i) \quad t_i(x), \quad x \in R^n\n\end{array} \tag{5}
$$

includes many of the most commonly used exponential families such as the binomial, $\overline{}$

$$
F^* = (a_1^*, \ldots, a_K^*, f_1^*, \ldots, f_K^*)
$$

$$
F^* \mathbf{A} , \qquad t_i(x)
$$

\n
$$
\mathbf{M} \qquad , \qquad P_i(x | \mathbf{f}_i^*) \qquad ;
$$

$$
P_i(x|\mathbf{f}_i^*) \leq U_i(x|\mathbf{f}_i^*) = w(x)(1^i)^{-c_1 - r(1/\mathbf{Z}_i(x))^c},\tag{6}
$$

^ZiðxÞ ¼ ^ðlⁱ maxÞ ni kx m i k and m ⁱ and lⁱ max are the mean vector and the maximum eigenvalue of the covariance matrix S ⁱ of Piðxjf ⁱ Þ; respectively. c1; c2; r and wðxÞ may depend on i implicitly. Furthermore, we assume that these lⁱ max are always bounded. We let n to be the least one among all these nⁱ and modify r such that these Uiðxjf ⁱ Þ are still the envelope functions of the component densities. As a result, we can let ^ZiðxÞ ¼ ^ðlⁱ maxÞ n kx m i k ; i ¼ 1; ... ; K,

$$
\mathbf{n} \qquad \qquad \blacksquare
$$

2.2. The EM algorithm and its asymptotic convergence rate

$$
f_1, ..., f_K
$$

\nML
\n
$$
{}^{(f)} = \{x^{(i)} : t = 1, ..., N\}
$$

\n
$$
I F = (a_1, ..., a_K, f_1, ..., f_K)
$$

\n
$$
L(F) = \sum_{t=1}^{N} P(x^{(t)} | F)
$$

\n
$$
E . (1), E M
$$

$$
\mathbf{a}_{i}^{+} = \frac{1}{N} \sum_{t=1}^{N} \frac{\mathbf{a}_{i} P_{i}(x^{(t)} | \mathbf{f}_{i})}{P(x^{(t)} | \mathbf{F})},\tag{7}
$$

$$
\mathbf{f}_{i}^{+} = \left\{ \sum_{t=1}^{N} t_{i}(x^{(t)}) \frac{\mathbf{a}_{i} P_{i}(x^{(t)} | \mathbf{f}_{i})}{P(x^{(t)} | \mathbf{F})} \right\} / \left\{ \sum_{t=1}^{N} \frac{\mathbf{a}_{i} P_{i}(x^{(t)} | \mathbf{f}_{i})}{P(x^{(t)} | \mathbf{F})} \right\},
$$
\n(8)

$$
L(F) \, |3,19 \,.\,M \qquad ,
$$

 \mathcal{E} regularity conditions, the EM is set \mathcal{E} z $L(F)$ |18. I ,
 \mathcal{S}_N , EM F^* (..., F^N EM and the true parameter F^* converges to the true parameter F^* correctly (i.e., **z** N \mathscr{S}_N is large, the EM algorithm converges to \mathbb{F}^N $N \rightarrow \infty$ $\mathbf{F}^N = \mathbf{F}^*$ in probability on \mathbf{z} F^* .

I ||18,
\n
$$
F^+ = G(F)
$$

\n $F^+ - F^N = G(F) - G(F^N) = G'(F^N)(F - F^N) + O(||F - F^N||^2)$ (9)
\nF O F^N, $G'(F)$ J $G(F)$ F^N $O(x)$

$$
x \to 0. \text{ B}
$$

\n
$$
E(G'(\mathbf{F}^*)) = I - Q(\mathbf{F}^*)R(\mathbf{F}^*),
$$

\n
$$
E(G'(\mathbf{F}^*)) = I - Q(\mathbf{F}^*)R(\mathbf{F}^*),
$$

$$
Q(\mathbf{F}^*) = diag(\mathbf{a}_1^*, \dots, \mathbf{a}_K^*, \mathbf{a}_1^{*-1} \mathbf{P}_1, \dots, \mathbf{a}_K^{*-1} \mathbf{P}_K)
$$
(10)

$$
P_{i} = \int_{R^{n}} [t_{i}(x) - f_{i}^{*}][t_{i}(x) - f_{i}^{*}] P_{i}(x|f_{i}^{*}) \text{ m}
$$

$$
R(\mathbf{F}^*) = \int_{R^n} V(x)V(x) \ P(x|\mathbf{F}^*) \ \mathbf{m} \tag{11}
$$

$$
V(x) = (b_1(x),...,b_K(x),a_1^*b_1(x)G_1(x),...,a_K^*b_K(x)G_K(x)),
$$

\n
$$
b_i(x) = P_i(x|f_i^*)/P(x|F^*),
$$

\n
$$
G_i(x) = P_i^{-1}[t_i(x) - f_i^*].
$$

Here $E(\cdot) = E_{\mathbf{F}^*}(\cdot)$. It follows from Eq. (9) \mathbb{E}^N is upper bounded by the norm kG^N $\|G'(F^N)\|$. B $\begin{array}{ccc} N & & \end{array}$ r and EM algorithm locally F^* : $r\leqslant \frac{1}{N\to\infty}\|G'(\mathrm{F}^N)\|=\bigg\|_{N\to\infty}\;G'(\mathrm{F}^N)$ $\begin{array}{c} \hline \end{array}$

$$
= ||E(G'(F^*))|| = ||I - Q(F^*)R(F^*)||.
$$
\n(12)\n
$$
I \qquad (12)
$$
\n
$$
EM
$$

 $\mathbf z$ \mathbf{z} .

3. The main result

3.1. The measures of the overlap

We revisit the measures used in [\[14\]](#page-23-0) for the overlap of component densities in a Gaussian mixture. We consider the following posterior densities for the mixture Eq. (1) with the true parameters F:

$$
h_i(x) = \frac{a_i^* P_i(x | f_i^*)}{\sum_{j=1}^K a_j^* P_j(x | f_j^*)}, \qquad i = 1, ..., K.
$$
 (13)

I E. (11)

$$
h_i(x) = \mathbf{a}_i^* \mathbf{b}_i(x).
$$
 (14)

$$
\mathbf{g}_{ij}(x) = (\mathbf{d}_{ij} - h_i(x))h_j(x) \qquad i, j = 1, \dots, K,
$$
\n
$$
\mathbf{d}_{ij} \qquad \mathbf{K} \qquad \dots \qquad ,
$$
\n
$$
\vdots
$$
\n(15)

$$
e_{ij}(\mathbf{F}^*) = \int_{R^n} |\mathbf{g}_{ij}(x)| P(x|\mathbf{F}^*) \mathbf{m}
$$

\n
$$
i, j = 1, 2, ..., K, \qquad e_{ij}(\mathbf{F}^*) \le 1 \qquad |\mathbf{g}_{ij}(x)| \le 1.
$$

\n
$$
\mathbf{F} \quad i \ne j, e_{ij}(\mathbf{F}^*)
$$

\n
$$
\begin{array}{ccc}\ni & j & P_i(x|\mathbf{f}^*) & P_j(x|\mathbf{f}^*) \\
x, h_i(x)h_j(x)\end{array}
$$

3.2. Regular conditions and lemmas

assumption prevents this degeneracy.

(1) Nondegenerate condition on the mixing proportions: mixing proportions satisfy the nondegenerate conditions satisfy the non-degenerate condition:

the exponential families to satisfy the following regular conditions: $\mathbf{f}_{\mathbf{f}}$

eigenvalues. The eigenvalues of all the covariance matrices satisfy

 $a_i^* \ge a$ $i = 1, ..., K,$ (16) a is a positive number of \mathbf{z} , corresponding component distribution will distribute α

(2) Uniform attenuating condition on the eigenvalues of the covariance matrices: S_i^* i be the covariance matrix of the item in l_{i1}, \ldots, l_{in}

degenerates to a mixture with a lower number of the mixing components. This

bl $(F^*) \leq l_{ij} \leq l(F^*)$ $i = 1, ..., K, k = 1, ..., n,$ (17) b is a positive number and loss defined to be the maximum eigenvalue of $\mathbf{l}(\mathbf{F}^*)$ $S_1^*, \ldots, S_K^*, \ldots,$ $l(F^*) = \int_{i,j} l_{ij}$ $B.$ $\qquad \qquad$ \mathbf{z} attenuate to \mathbf{z} . It $E(17)$ that the condition numbers of the K covariance matrices are K covariance matrices are K covariance matrices are K $\,$, $\,$, $\,$, $\,$, $1 \leqslant k(S_i^*) \leqslant B'$ $i = 1, ..., K,$ $k(S_i^*)$ \mathbf{S}_i^* **S** $\stackrel{*}{_{i}}\qquad B'$ is a positive number.

(3) Regular condition on the mean vectors: $m_{\tilde{t}}^*$

$$
mD \t(F^*) \le D \t(F^*) \le ||m_i^* - m_j^*|| \le D \t(F^*) \t i \ne j, \t (18)
$$

\n
$$
D \t(F^*) = \t i \ne j ||m_i^* - m_j^*||, \t D \t(F^*) = \t i \ne j ||m_i^* - m_j^*||, \t m
$$

\n
$$
M \t , \t m_i^*, m_j^* \t , \t j, I \t m_i^* \t ... \t m_l^*, \t ... \t m_l^* \t m_j^*|| \ge T \t i \ne j, I \t , \t m_l^* \t ... \t m_l^* \t ... \t m_l^* \t m_l^* ||P_i^{-1}||, \t P_i \t E . (10).
$$

\n
$$
P_i(x|f_i^*)
$$

\n
$$
P_i =
$$

\n(10)

$$
Z(F^*) \to 0.
$$
\n
$$
Z(F^*) \to 0
$$
\n
$$
Z(F^
$$

$$
e_{ij}(\mathbf{F}^*) \leq e(\mathbf{F}^*) \leq f(\mathbf{Z}(\mathbf{F}^*)) \qquad i \neq j. \tag{22}
$$

$$
F \qquad , \qquad \qquad (\qquad A \qquad A \qquad).
$$

Lemma 1. Suppose that a mixture of K densities from the bell sheltered exponential families of the parameter F^* satisfies Conditions (1)–(3). As $Z(F^*)$ tends to zero, we have

- () $Z(F^*)$, $Z_i(m_j^*)$ and $Z_j(m_i^*)$ are the equivalent infinitesimals.
- () For $i \neq j$, we have

$$
||m_i^*|| \le T'||m_i^* - m_j^*||,
$$
\n(23)

where T' is a positive number.

(e) For any two nonnegative numbers with $p + q > 0$, we have

$$
||m_i^* - m_j^*||^p (1^i)^{-nq} \le O(Z^{-p \vee q}(F^*)),
$$
\n(24)

where $p \vee q = \{p, q\}.$

Lemma 2. Suppose that a mixture of K densities from the bell sheltered exponential families of the parameter F^* satisfies Conditions (1)–(3). As $Z(F^*)$ tends to zero, we have for each i

$$
\|P_i\| \leqslant c \|m_i^* - m_j^*\|^p,\tag{25}
$$

where $j \neq i$, c and p are some positive numbers.

$$
() \t E(||t_i(X) - f_i^*||^2) \leq u M_i^q(F^*), \t (26)
$$

where $M_i(\mathbf{F}^*) = \int_{\mathbf{F}} \|\mathbf{m}_i^* - \mathbf{m}_j^*\|$, **u** and q are some positive numbers.

Lemma 3. Suppose that a mixture of K densities from the bell sheltered exponential families of the parameter F^* satisfies Conditions (1) (3) and $Z(F^*) \rightarrow 0$ as an infinitesimal, we have

$$
f^{\mathfrak{e}}(\mathcal{Z}(\mathbf{F}^*)) = o(\mathcal{Z}^p(\mathbf{F}^*)),\tag{27}
$$

where $e>0$, p is any positive number and $o(x)$ means that it is a higher order infinitesimal as $x \to 0$.

characteristic of the densities from the bell sheltered exponential families.

 $\mathcal{S}_{\mathcal{A}}$, we are ready to give our main theorem.

 \overline{E} appear in the main theorem. Especially, the main theorem. Espe $e(F^*)$ and $Z(F^*)$ and by Lemma 3 and the flux reflects the flux reflection by $L = 3$ and the flux reflection F

3.3. The main theorem

Theorem 1. Given a mixture of K densities from the bell sheltered exponential families of the parameter F^* that satisfies Conditions (1)–(4), as $e(F^*)$ tends to zero as an infinitesimal, we have

$$
r \leq \|E(G'(\mathbf{F}^*))\| = o(\ ^{0.5-\mathbf{e}}(\mathbf{F}^*)),\tag{28}
$$

where e is an arbitrarily small positive number.

 $A \qquad \qquad , \qquad \qquad , \qquad \qquad ,$ becomes small, $e(F^*) \to 0$, $||E(G'(F^*))||$ $0.5 - e(F^*)$. (F^*) . (F^*) to zero, the asymptotic to F^* EM and F^* $^{0.5-e}$ (F^{*}). (F^*) . , $e(F^*)$ N , EM and z in other approximately zero. In other approximately zero. In other approximately zero. In other approximately z in z \mathcal{L}_{N} and the EM algorithm in this case has a \mathcal{L}_{N} $\mathbf M$ is follows from the theorem that the asymptotic convergence convergence convergence convergence convergence convergence $\mathbf M$ \mathcal{I} with the overlap measure as it tends to zero. This t measure in the overlap measure in the mixture $\mathcal S$ EM increases greatly may provide a theoretic basis for the study of the study convergence rate EM in the cases of finite overlap and data. $\hspace{.1cm}0\hspace{1.2cm}$, the theorem has also provided a new mathematical provided a $t_{\rm eff}$ that the rate of the FM algorithm is determined by the fraction is determined by the fractio of missing-data information. Actually, the measure of overlap and \mathcal{A} densities is equivalent to the fraction of missing-data information in the mixture. $\mathcal I$, the component density of each sample data tends to zero, the component data sample data samp becomes very clear. That is, the fraction of missing-data information \mathbf{z} information \mathbf{z} T , the overlap measure can be considered as the fraction of missing-data be cons information in the mixture. In this way, the third proved that the EM EM algorithm tends to converge superlinearly as the fraction of missing-data σ \boldsymbol{z} information tends to be zero. Moreover, the provides another provides anothe for the acceleration methods like the acceleration methods like the ''working parameter'' methods like the '' $[16$ and P -EM algorithm $[13]$, the concept that the EM algorithm will have a fast rate of convergence if the fraction of missing-data fraction of \mathcal{A} information is small.

Proof of Theorem 1.
\n
$$
Q(F^*)R(F^*)
$$
 , $Q(F^*)R(F^*)$.
\n $Q(F^*)R(F^*)$, $Q(F^*)R(F^*)$.
\n $Q(F^*)R(F^*)$ = $diag[diag[\mathcal{A} \mid A^{*-1}P_1, ..., A_K^{*-1}P_K]$
\n $\times \begin{pmatrix} R_{b,b} & R_{b,G_1} & \cdots & R_{b,G_K} \\ R_{G_1,b} & R_{G_1,G_1} & \cdots & R_{G_K,G_K} \\ \vdots & \vdots & \ddots & \vdots \\ R_{G_K,b} & R_{G_K,G_1} & \cdots & R_{G_K,G_K} \end{pmatrix}$
\n $= \begin{pmatrix} diag[\mathcal{A} \mid R_{b,b} & diag[\mathcal{A} \mid R_{b,G_1} & \cdots & diag[\mathcal{A} \mid R_{b,G_K} \\ a_1^{*-1}P_1R_{G_1,b} & a_1^{*-1}P_1R_{G_1,G_1} & \cdots & a_1^{*-1}P_1R_{G_1,G_K} \\ \vdots & \vdots & \ddots & \vdots \\ a_K^{*-1}P_KR_{G_K,b} & a_K^{*-1}P_KR_{G_K,G_1} & \cdots & a_K^{*-1}P_KR_{G_K,G_K} \end{pmatrix}$,
\n $R(F^*)$
\n $V(x) = [b_1(x), ..., b_K(x)]$ $\mathcal{A} = [a_1^*, ..., a_K^*]$.
\n $R(K^*)$
\n $V(x) = [b_1(x), ..., b_K(x)]$ $\mathcal{A} = [a_1^*, ..., a_K^*]$.
\n $V(x)$
\n $V(x) = [b_1(x), ..., b_K(x)]$ $\mathcal{B} = [a_1^*, ..., a_K^*]$
\n $V(x)$
\n $V(x) = a_1^*b_1(x),$
\n $\int_{R^m} b_i(x)b_j(x)P(x|F^*)$ $\mathbf{m} = \frac{1}{a_1^*} \frac{1}{a_1^*j^2} e_{ij}(F^*)$ $i \neq j$,
\n $\int_{R^m} b_i(x)P(x|F^*)$ \math

1 a_j^* $e_{ij}(\mathbf{F}^*) \leq \frac{1}{2}$ $\frac{1}{a}e_{ij}(F^*)=o(\sqrt{0.5-e}(F^*)),$ $diag[\mathcal{A} \;]R_{b,b} = I_K + o$

 $\mathcal{S}_{\mathcal{S}}$, we consider the above first matrix $\mathcal{S}_{\mathcal{S}}$. It follows from the above first matrix from the above first matrix $\mathcal{S}_{\mathcal{S}}$ C \mathbf{z} in $\left| \mathbf{g}_{ii}(x) \right| \leq 1$ $|E(h_j(X)(h_i(X) - d_{ij})(t_{i,k}(X) - f_{i,k}^*))|$ $\leq E(|h_j(X)(h_i(X) - d_{ij})| |(t_{i,k}(X) - f_{i,k}^*)|)$ $\leq E^{1/2}(\mathsf{g}_{ij}^2(X))E^{1/2}((t_{i,k}(X)-\mathsf{f}_{i,k}^*)^2)$ $\leq E^{1/2}(|g_{ij}(X))E^{1/2}((t_{i,k}(X)-f_{i,k}^*)^2)$ \leq $\sqrt{e_{ij}(\mathbf{F}^*)}E^{1/2}((t_{i,k}(X)-\mathbf{f}^*_{i,k})^2).$ A $L_{i/2} = 2, E(\|t_i(X) - \mathbf{f}_i^*\|^2 | \mathbf{F}^*)$ u M_i^q $\mathfrak{u}M_i^q(\mathrm{F}^*).$
 $\sqrt{\mathfrak{u}M_i^q(\mathrm{F}^*)}.$
 $\sqrt{\mathfrak{u}M_i^q(\mathrm{F}^*)}.$ $E^{1/2}((t_{i,k}(X)-{\bf f}^{\ast}_{i,k})^2$ $\overline{}$, $E(diag[\mathscr{A}]\mathbf{a}_i^*\mathbf{b}_i(X)\mathbf{b}(X)(t_i(X) - \mathbf{f}_i^*) = O(M_i^{q/2}(\mathbf{F}^*)e^{0.5}(\mathbf{F}^*)).$ A Lemmas 1 and 3, $M_i^{q/2}$ (F^{*}) $^{0.5}$ (F^{*}) e (F^{*}) $Z(F^*)$ **z** . I $||E(diag[\mathcal{A} \;]a_i^*b_i(X)b(X)(t_i(X) - f_i^*)|| = O(M_i^{q/2}(F^*)^{0.5}(F^*)).$ $M \qquad ,$ $\|\text{diag}[\mathcal{A} \mid] R_{b,G_i} \| \leq \|E(\text{diag}[\mathcal{A} \mid]a_i^* b_i(X)b(X)(t_i(X) - f_i^*) \| \| \|P_i^{-1}\|$ $||P_i^{-1}|| = ||I(f_i^*)|| \le O(||m_i^*||^{t_1} (1^i))^{-t_2})$ $C \t(4).$ $\|\text{diag}[\mathscr{A}]\|R_{\mathrm{b},\mathrm{G}_i}\| \leq \mathbf{u} \|m_i^* - m_j^*\|^{q_1} (\mathbf{1}^i)^{-q_2} \sim 0.5(\mathrm{F}^*),$ $q_1 = (q/2) + t_1, q_2 = t_2,$ u $\qquad \qquad \text{L} \qquad \qquad \text{1} \qquad \text{3}$ $\text{diag}[\mathscr{A} \;] R_{\mathbf{b}, \mathbf{G}_i} \parallel \leqslant O(\mathbf{Z}^{-q_1 \vee q_2}(\mathbf{F}^*)) e^{0.5}(\mathbf{F}^*) = o(e^{0.5 - \mathbf{e}}(\mathbf{F}^*)).$ $\mathbf B$ the properties of matrix $\mathbf B$ $diag[\mathcal{A} \;]R_{\mathbf{b},\mathbf{G}_i} = o(\ ^{0.5-\mathbf{e}}(\mathbf{F}^*)).$ () The computation of $a_i^{*-1}P_iR_{G_i,b}$ $(i = 1, ..., K)$: A (1) Letter $2, a_i^{*-1} \|P_i\|$ is upper bounded by $(1/a)c \|m_i^* - m_j^*\|^p$, $j \neq i$, c and p $R_{\text{G}_i,\text{b}} = R_{\text{b},\text{G}_i},$ (b) in $a_i^{*-1} P_i R_{G_i, b} = o(^{0.5-e}(F^*)).$

() The computation of $\mathbf{a}_{i}^{*-1} \mathbf{P}_{i} R_{\mathbf{G}_{i},\mathbf{G}_{i}}$ $(i = 1, ..., K)$: B $V(x)$,

$$
a_i^{*-1}P_iR_{G_i,G_i} = a_i^{*-1}P_iE(h_i^2(X)G_i(X)G_i(X))
$$

= $a_i^{*-1}E(h_i^2(X)(t_i(X) - f_i^*)(t_i(X) - f_i^*))P_i^{-1}$
= $I_{d_i} + a_i^{*-1}E(h_i(X)(h_i(X) - 1)(t_i(X) - f_i^*)(t_i(X) - f_i^*))P_i^{-1}$,
:

 $P_i E(h_i(X) G_i(X) G_i(X)) = a_i^* I_{d_i}.$ F, $E(\|t_i(X) - \mathbf{f}_i^*\|^2 | \mathbf{F}^*)$ u M_i^q $u M_i^q(\mathrm{F}^*)$ a_i^{*-1} is upper bounded, in a similar way as above, we can prove i $a_i^{*-1} E(h_i(X)(h_i(X) - 1)(t_i(X) - f_i^*)(t_i(X) - f_i^*)) P_i^{-1} = o(\sqrt{0.5^{-e}}(F^*))$ $\,$, $\,$ $a_i^{*-1}P_iR_{G_i,G_i} = I_{d_i} + o(^{-0.5-e}(F^*)).$ () The computation of $a_i^{*-1}P_iR_{G_i,G_j}$ $(j \neq i)$: B $V(x)$, $\mathbf{a}_{i}^{*-1} \mathbf{P}_{i} R_{\mathbf{G}_{i},\mathbf{G}_{j}} = \mathbf{a}_{i}^{*-1} E(\mathbf{a}_{i}^{*} \mathbf{b}_{i}(X) \mathbf{a}_{j}^{*} \mathbf{b}_{j}(X)(t_{i}(X) - \mathbf{f}_{i}^{*})(t_{j}(X) - \mathbf{f}_{j}^{*}) \mathbf{P}_{j}^{-1}$ $= a_i^{*-1} E(h_i(X)h_j(X)(t_i(X) - f_i^*)(t_j(X) - f_j^*)) P_j^{-1}.$ $\left(\right)$, $a_i^{*-1}P_iR_{G_i,G_j} = o(^{0.5-e}(F^*)).$ $()$ (b), we obtain: $Q(F^*)R(F^*) = I + o(^{0.5-e}).$

 E . (12), $r \leq ||I - Q(F^*)R(F^*)|| = o(^{0.5 - e}(F^*)$. \Box

4. A typical class: Gaussian mixtures

multivariate normal family.

 EM

 G families. As example, $P_i(x|m_i, S_i)$ by Eq. (2) $y_i = (S_i^{-1}m_i, S_i^{-1})$ $t_i(x) = (x, -\frac{1}{2})$ $\frac{1}{2}xx$)., \int_{t}^{1} y_i, \int_{t}^{1} y_i, $(m_i, -\frac{1}{2}(S_i + m_i m_i))$ (m_i, S_i)

 N_{eff} , which indicates that Gaussian densities are Gaus bell-sheltered if the condition numbers of their covariance matrices are upper

bounded.

Lemma 4. Suppose that $P_i(x|f_i^*) = P_i(x|m_i^*, S_i^*)$ is a Gaussian distribution with the mean m_i^* and the covariance matrix S_i^* , and that the condition number of S_i^* , i.e., $k(S_i^*)$, is upper bounded by B'. We have that $P_i(x|\hat{f}_i^*)$ is bell-sheltered, i.e.,

$$
P_i(x|\mathbf{f}_i^*) = P_i(x|m_i^*, \mathbf{S}_i^*) \le b \frac{1}{(\mathbf{1}^i - \mathbf{1}^i)^{n/2}} \quad \text{and} \quad \mathbf{I} = \mathbf{1}^i \mathbf{1}^i \mathbf{1}^i \mathbf{1}^i \tag{29}
$$

where *b* is a positive number.

Proof. B
\n
$$
y = U_i(x - m_i^*)
$$
\n
$$
P(y|1^i) = \frac{1}{(2pl^i)^{n/2}} e^{-(1/21^i) ||y||^2},
$$

$$
P_i(x|m_i^*, \mathbf{S}_i^*) \leq B^{n/2} P(y|\mathbf{l}^i),
$$

\n
$$
k(\mathbf{S}_i^*) \leq B'. M, \qquad ||y|| = ||x - m_i^*||,
$$

\n
$$
P_i(x|m_i^*, \mathbf{S}_i^*) \leq b \frac{1}{(\mathbf{l}^i)^{n/2}} \quad \text{and} \quad \text{and}
$$

M, $P_i(x|f_i^*) = P_i(x|m_i^*, S_i^*)$, $t_i(x)$

 f_i^*

$$
t_{i}(x) = \begin{cases} x & m_{i}^{*}, \\ -\frac{1}{2}xx & -\frac{1}{2}(S_{i}^{*} + m_{i}^{*}(m_{i}^{*})) .\end{cases}
$$

\n
$$
t_{i}(x) = \begin{cases} x & m_{i}^{*}, \\ -\frac{1}{2}xx & -\frac{1}{2}(S_{i}^{*} + m_{i}^{*}(m_{i}^{*})) .\end{cases}
$$

\n
$$
t_{i}(x) \qquad x_{1},...,x_{n}.
$$

\n
$$
F^{*} \qquad (1) (3)
$$

\n
$$
f_{i}^{*} = [(m_{i}^{*}) , vec[S_{i}^{*}]], \qquad S_{i}^{*} = -\frac{1}{2}(S_{i}^{*} + m_{i}^{*}(m_{i}^{*})) , \qquad \hat{f}_{i}^{*} = [(m_{i}^{*}) , vec[S_{i}^{*}]].
$$

Lemma 5. Suppose that $P_i(x|f_i^*) = P_i(x|m_i^*, S_i^*)$ is a Gaussian density and $k(S_i^*)$ is upper bounded. As l^i tends to zero, we have

$$
||I(\mathbf{f}_i^*)|| = O((1^i -)^{-t}), \tag{30}
$$

where t is a positive number.

Proof. B_y

 $\overline{}$

$$
\frac{\partial P_i(x|m_i^*, \mathbf{S}_i^*)}{\partial m_i^*} = (x - m_i^*) \mathbf{S}_i^* P_i(x|m_i^*, \mathbf{S}_i^*),
$$
\n(31)\n
$$
\frac{\partial P_i(x|m_i^*, \mathbf{S}_i^*)}{\partial \mathbf{S}_i^*} = -\frac{1}{2} (\mathbf{S}_i^{*-1} - \mathbf{S}_i^{*-1} (x - m_i^*) (x - m_i^*) \ \mathbf{S}_i^{*-1}) P_i(x|m_i^*, \mathbf{S}_i^*).
$$
\n(32)

A cording to the definition of the definition of the $\mathbb H$ information matrix, we also have $\mathbb H$

$$
I(\mathbf{f}_{i}^{*}) = E_{\mathbf{f}_{i}^{*}} \left(\left(\frac{\partial P_{i}(X|\mathbf{f}_{i}^{*})}{\partial \mathbf{f}_{i}^{*}} \right) \left(\frac{\partial P_{i}(X|\mathbf{f}_{i}^{*})}{\partial \mathbf{f}_{i}^{*}} \right) \right)
$$

\n
$$
= E_{\mathbf{f}_{i}^{*}} \left(\frac{\partial(\hat{\mathbf{f}}_{i}^{*})}{\partial \mathbf{f}_{i}^{*}} \left(\frac{\partial P_{i}(X|\mathbf{f}_{i}^{*})}{\partial \hat{\mathbf{f}}_{i}^{*}} \right) \left(\frac{\partial P_{i}(X|\mathbf{f}_{i}^{*})}{\partial \hat{\mathbf{f}}_{i}^{*}} \right) \left(\frac{\partial(\hat{\mathbf{f}}_{i}^{*})}{\partial \mathbf{f}_{i}^{*}} \right) \right)
$$

\n
$$
= \frac{\partial(\hat{\mathbf{f}}_{i}^{*})}{\partial \mathbf{f}_{i}^{*}} I(\hat{\mathbf{f}}_{i}^{*}) \left(\frac{\partial(\hat{\mathbf{f}}_{i}^{*})}{\partial \mathbf{f}_{i}^{*}} \right) ,
$$

$$
I(\hat{\mathbf{f}}_i^*) = E_{\mathbf{f}_i^*} \left(\left(\frac{\partial P_i(X|\mathbf{f}_i^*)}{\partial \hat{\mathbf{f}}_i^*} \right) \left(\frac{\partial P_i(X|\mathbf{f}_i^*)}{\partial \hat{\mathbf{f}}_i^*} \right) \right).
$$

\nI. E. (31)
\n
$$
P_i^3(x|m_i^*, S_i^*)
$$
\n(32)
\n
$$
P_i(x|m_i^*, \frac{1}{3}S_i^*)
$$
\n
$$
I(\hat{\mathbf{f}}_i^*)
$$
\n
$$
|S_i^*|
$$
\nG

$$
I(\hat{f}_{i}^{*}) = E_{(m_{i}^{*}, (1/3)S_{i}^{*})}(G(X, f_{i}^{*})),
$$
\n
$$
G(x, f_{i}^{*}) \t x - m_{i}^{*} \t S_{i}^{*}.
$$
\n
$$
y = x - m_{i}^{*},
$$
\n
$$
I(\hat{f}_{i}^{*}) = E_{(0,(1/3)S_{i}^{*})}(G(Y, S_{i}^{*})),
$$
\n
$$
G(y, S_{i}^{*}) \t S_{i}^{*-1}
$$
\n
$$
y_{1},..., y_{n}. I \t S_{i}^{*-1}
$$
\n
$$
S_{i}^{*-1} = |S_{i}^{*}|^{-1} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d_{i}} \\ a_{21} & a_{22} & \cdots & a_{2d_{i}} \\ \vdots & \vdots & \ddots & \vdots \\ a_{d_{i1}} & a_{d_{i2}} & \cdots & a_{d_{i}d_{i}} \end{pmatrix},
$$
\n
$$
a_{kl}
$$
\n
$$
g_{pq}(y, S_{i}^{*}) \t S_{kl}^{*}
$$
\n
$$
S_{kl}^{*}
$$
\n
$$
g_{pq}^{*}
$$
\n
$$
S_{kl}^{*}
$$
\n
$$
G_{kl}^{*}
$$
\n
$$
g_{pq}^{*}
$$
\n
$$
S_{kl}^{*}
$$
\n
$$
S_{kl}^{*}
$$

 $\begin{array}{ll} *i & S_i^*, \\ k l & \end{array}$

G
\n
$$
(1^{i})^{t}g_{pq}(y, S_{i}^{*})
$$
\n
$$
1^{i} < B, \qquad E_{(0,1/3S_{i}^{*})}((1^{i})^{t}g_{pq}(Y, S_{i}^{*}))
$$
\n
$$
||I(\hat{f}_{i}^{*})|| = ||(1^{i})^{-t}(1^{i})^{t}I(f_{i}^{*})||
$$
\n
$$
= (1^{i})^{-t}||(1^{i})^{t}I(f_{i}^{*})||
$$
\n
$$
\leq o(1^{i})^{-t},
$$
\n
$$
B
$$
\n
$$
||I(f_{i}^{*})|| \leq ||\frac{\partial(\hat{f}_{i}^{*})}{\partial f_{i}^{*}}|| ||I(\hat{f}_{i}^{*})|| || \left(\frac{\partial(\hat{f}_{i}^{*})}{\partial f_{i}^{*}}\right) || = 4||I(\hat{f}_{i}^{*})||,
$$
\n
$$
||\partial(\hat{f}_{i}^{*}) / \partial f_{i}^{*}|| = ||(\partial(\hat{f}_{i}^{*}) / \partial f_{i}^{*})|| = 2,
$$
\n
$$
E . (30)
$$
\n
$$
EM
$$
\n
$$
(1) (3)
$$

Theorem 2. Given a Gaussian mixture of K densities of the parameter F^* that satisfies conditions (1)–(3), as $e(F^*)$ tends to zero as an infinitesimal, we have

$$
||G'(\mathbf{F}^*)|| = o(^{0.5-\mathbf{e}}(\mathbf{F}^*)), \tag{33}
$$

where e is an arbitrarily small positive number.

In other words, Theorem 1 applies to Gaussian mixture when only conditions to Gaussian mixture when only conditions G (1)–(3) \blacksquare

5. Conclusions

Acknowledgements

This work was supported by a Grant from the Research Grant Councilof the Hong Kong SAR (Project no. CUHK4225/04E) and a Grant from the Natural Science Foundation of China (Project no. 60071004).

Appendix

Proof of Lemma 1.
\n
$$
Z(F^*) = i_{\neq j} Z_i(m_j^*) = Z_i(m_j^*). A
$$
\n
$$
Z(F^*) = \sum_{i \neq j} Z_i(m_j^*) = Z_i(m_j^*). A
$$
\n
$$
a_1(1^i)^n \leq (1^i)^n \leq a_2(1^i)^n,
$$
\n
$$
b_1(1^i)^n \leq (1^i)^n \leq b_2(1^i)^n,
$$
\n
$$
c_1 || m_i^* - m_j^* || \leq || m_i^* - m_j^* || \leq c_2 || m_i^* - m_j^* ||.
$$
\n
$$
C = E . (34) (35) E . (36),
$$
\n
$$
d'_1, d'_2, b'_1, b'_2
$$
\n
$$
d'_1 Z(F^*) \leq Z_i(m_j^*) \leq d'_2 Z(F^*),
$$
\n
$$
b'_1 Z_i(m_j^*) \leq Z_j(m_j^*) \leq d'_2 Z(F^*).
$$
\n
$$
Z(F^*) \geq 2m_j^* ||
$$
\n
$$
R(F^*) \geq 0, E . (23)
$$
\n
$$
|| m_i^* || = || m_i^* - m_j^* ||
$$
\n
$$
|| m_i^* || = || m_j^* ||
$$
\n
$$
|| m_i^* || = || m_j^* ||
$$
\n
$$
|| m_i^* || = || m_j^* ||
$$
\n
$$
|| m_i^* || = || m_j^* ||
$$
\n
$$
|| m_i^* || = || m_i^* ||
$$
\n
$$
|| m_i^* || = || m_i^* ||
$$
\n
$$
|| m_i^* || = || m_i^* ||
$$
\n
$$
|| m_i^* || = || m_i^* ||
$$
\n
$$
|| m_i^* || = || m_i^* ||
$$
\n
$$
|| m_i^* || = || m_i^* ||
$$
\n
$$
|| m_i^* || = || m_i^* ||
$$
\n
$$
|| m_i^* || = || m_i^* ||
$$
\n
$$
|| m_i^* || = || m_i^* ||^2 (1^i)^{-1} || = || m_i^* ||^2 (1^i)^{-1} ||
$$
\n
$$
|| m_i^*
$$

 $t_i(x) = P_0 + P_1x + P_2x^2 + \cdots + P_kx^k,$

where the following expression is into the following expression:

$$
k \ge 0
$$
, P_i $d_i \times n^i$, x^i
\n $x_{1, x_2, ..., x_n}$. I
\n
$$
\begin{array}{ccc}\nx_{j_1}x_{j_2}...x_{j_i} & x^i \\
B & \n\end{array}
$$
\n $||x^i|| \le \sqrt{n}||x||^i$ $i = 0, 1, ..., k$.

$$
t_i(x) = t_i(x - m_i^* + m_i^*)
$$

= $P'_0 + P'_1(x - m_i^*) + P'_2(x - m_i^*)^2 + \dots + P'_k(x - m_i^*)^k,$ (38)
 P'_i $d_i \times n^i$, $m_{i1}^*, \dots, m_{in}^*$.

$$
f_i^* = E_{f_i^*}(t_i(X)) = P'_0 + E_{f_i^*}(P'_1(X - m_i^*)) + \dots + E_{f_i^*}(P'_k(X - m^*)^k)
$$
(39)

$$
E_{f_i^*}(P'_1(X - m_i^*)) = P'_1E_{f_i^*}(X - m_i^*) = 0,
$$

$$
t_i(X) - f_i^* = \sum_{j=1}^k [P'_j(X - m_i^*)^j - E_{f_i^*}(P'_j(X - m_i^*)^j)].
$$

 $N,$

$$
E_{f_i^*}(\|t_i(X) - f_i^*\|^2) = E_{f_i^*}(\|(t_i(X) - f_i^*) (t_i(X) - f_i^*)\|)
$$

\n
$$
= E_{f_i^*} \Biggl(\Biggl\| \sum_{|j_1|=1, j_2=1}^k [P'_{j_1}(X - m_i^*)^{j_1} - E_{f_i^*}(P'_{j_1}(X - m_i^*)^{j_1})]
$$

\n
$$
\times [P'_{j_2}(X - m_i^*)^{j_2} - E_{f_i^*}(P'_{j_2}(X - m_i^*)^{j_2})] \Biggr\| \Biggr)
$$

\n
$$
\leq \sum_{j_1=1, j_2=1}^k E_{f_i^*}(\|[P'_{j_1}(X - m_i^*)^{j_1} - E(P'_{j_1}(X - m_i^*)^{j_1})]\|
$$

\n
$$
\times \|[P'_{j_2}(X - m_i^*)^{j_2} - E_{f_i^*}(P'_{j_2}(X - m_i^*)^{j_2})]\|)
$$

\n
$$
\leq \sum_{j_1=1, j_2=1}^k E_{f_i^*}^{1/2}(\|P'_{j_1}(X - m_i^*)^{j_1} - E_{f_i^*}(P'_{j_1}(X - m_i^*)^{j_1})\|^2)
$$

\n
$$
\times E_{f_i^*}^{1/2} \|P'_{j_2}(X - m_i^*)^{j_2} - E_{f_i^*}(P'_{j_2}(X - m_i^*)^{j_2} |f_i^*)\|^2).
$$
 (40)

 $P \t\t$,

$$
E_{\mathbf{f}_i^*}(\|P'_{j_1}(X - m_i^*)^{j_1} - E_{\mathbf{f}_i^*}(P'_{j_1}(X - m_i^*)^{j_1})\|^2)
$$

\n
$$
= E_{\mathbf{f}_i^*}(\|P'_{j_1}(X - m_i^*)^{j_1}\|^2) - \|E_{\mathbf{f}_i^*}(\|P'_{j_1}(X - m_i^*)^{j_1})\|^2
$$

\n
$$
\leq E_{\mathbf{f}_i^*}(\|P'_{j_1}(X - m_i^*)^{j_1}\|^2) \leq \sqrt{n}E_{\mathbf{f}_i^*}(\|P'_{j_1}\|^2 \|X - m_i^*\|^{2j_1})
$$

\n
$$
= \sqrt{n}||P'_{j_1}||^2 E_{\mathbf{f}_i^*}(\|X - m_i^*\|^{2j_1}).
$$
\n(41)

B
$$
P_i(x|\mathbf{f}_i^*) \le U_i(x|\mathbf{f}_i^*),
$$

\n
$$
E_{\mathbf{f}_i^*}(\|X - m_i^*\|^{2j_1}) \le \int \|x - m_i^*\|^{2j_1} U_i(x|\mathbf{f}_i^*) \, x
$$
\n
$$
= \int \|y\|^{2j_1} w(y + m_i^*) (1^i)^{-c_1 - r(1/(1^i)^{nc_2})\|y\|^{c_2}} y,
$$
\n(42)\n
$$
y = x - m_i^*, \qquad w(x)
$$

$$
w(y + m_i^*) \leq w_0 + w_1 \|y\| + \dots + w_{k'} \|y\|^{k'},
$$
\n
$$
k' , w_0, w_1, \dots, w_{k'}
$$
\n
$$
\|m_i^*\|, \dots,
$$
\n
$$
w_i = w_0^i + w_1^i \|m_i^*\| + \dots + w_{c_i}^i \|m_i^*\|^{c_i} \qquad i = 0, 1, \dots, k',
$$
\n
$$
w_0^i, w_1^i, \dots, w_{c_i}^i, \dots, w_{c_i}^i, \dots, w_{c_i}^i, \dots, w_{c_i}^i
$$
\n
$$
B \quad L \qquad 1,
$$
\n
$$
w_i \leq v_0^i + v_1^i \|m_i^* - m_j^*\| + \dots + v_{c_i}^i \|m_i^* - m_j^*\|^{c_i} \qquad i = 0, 1, \dots, k',
$$
\n
$$
v_0^i, v_1^i, \dots, v_{c_i}^i, \dots, v_{c_i}^i, \dots, w_{c_i}^i, \dots, w_{c_i}^i
$$
\n
$$
E \quad (42),
$$
\n
$$
(43)
$$

$$
E_{f_i^*}(\|X - m_i^*\|^{2j_1}) \leq \sum_{l=0}^k w_l (1^i)^{-c_1} \int \|y\|^{2j_1 + l} - r(1/(1^i)^{n c_2}) \|y\|^{c_2} y
$$

\n
$$
= \sum_{l=0}^k w_l (1^i)^{-c_1 + n(2j_1 + l + 1)} \int \|u\|^{2j_1 + l} - r \|u\|^{c_2} u,
$$

\n
$$
u = y/(1^i)^n \cdot C, \quad \int \|u\|^{2j_1 + l} - r \|u\|^{c_2} u
$$

\n
$$
E_{f_i^*}(\|X - m_i^*\|^{2j_1})
$$

\n
$$
\begin{array}{c} j_1 \\ \vdots \\ j_l \end{array}
$$

\n
$$
P'_{j_1} \qquad m_i^* \| \dots m_i^* \dots m_m^* \cdot \|P'_{j_l} \|
$$

\n
$$
\begin{array}{c} m_i^* - m_j^* \| \dots \\ m_i^* - m_j^* \| \dots \\ m_i^* - m_j^* \| \dots \end{array}
$$

\n
$$
E_{f_i^*}(\|P'_{j_1}(X - m_i^*)^{j_1} - E_{f_i^*}(P'_{j_1}(X - m_i^*)^{j_1})\|^{2})
$$

\n
$$
E_{f_i^*}(\|P'_{j_1}(X - m_i^*)^{j_1} - E_{f_i^*}(P'_{j_1}(X - m_i^*)^{j_1})\|^{2})
$$

\n
$$
E_{f_i^*}(\|P'_{j_1}(X - m_i^*)^{j_1} - E_{f_i^*}(P'_{j_1}(X - m_i^*)^{j_1})\|^{2}) \leq c_{j_1} \|m_i^* - m_j^*\|^{p_{j_1}},
$$

\n
$$
C_{j_1} \qquad p_{j_1} \qquad E \quad (40),
$$

\n
$$
E_{f_i^*}(\|t_i(X) - f_i^*\|^{2}) \leq c \|m_i^* - m_j^*\|^{p},
$$

\n
$$
E \quad (37), ()
$$

the assumptions.

A (),
$$
j \neq i
$$
, $f'_j = E_{f_j^*}(t_i(X))$
\n
$$
E_{f_j^*}(\|t_i(X) - f_i^*\|^2) \leq E_{f_j^*}((\|t_i(X) - f_j'\| + \|f_j' - f_i^*\|)^2)
$$
\n
$$
= E_{f_j^*}(\|t_i(X) - f_j'\|^2 + 2\|t_i(X) - f_j'\|\|f_j' - f_i^*\| + \|f_j' - f_i^*\|^2)
$$
\n
$$
\leq E_{f_j^*}(2\|t_i(X) - f_j'\|^2 + 2\|f_j' - f_i^*\|^2)
$$
\n
$$
= 2E_{f_j^*}(\|t_i(X) - f_j'\|^2) + 2\|f_i^* - f_j'\|^2. \tag{48}
$$

$$
\cdot \mathbf{I}
$$

$$
E_{f_j^*}(\|t_i(X) - f_j'\|^2) \leq c_1 \|m_i^* - m_j^*\|^{p_1},
$$

\n
$$
c_1 \quad p_1 \quad \dots \quad M \quad ,
$$

\n
$$
\|f_i^* - f_j'\| \leq \|f_i^*\| + \|f_j'\|.
$$
\n(49)

 \blacksquare , we can prove that \blacksquare

 $B E . (38),$ $\| \mathbf{f}'_j \|$ and $\| \mathbf{f}'_j \|$ $c_2 \| m_i^* - m_j^*$ C_2 and p_2 are some positive numbers. Therefore, p_2 **b** . (48), $E_{f_j^*}(\|t_i(X) - f_i^*$ $\|m_i^*\|^{2})$ is upper bounded by a polynomial of $\|m_i^*\|$ $m_j^* \| .$ $\qquad \quad \| m_i^{*'} - m_j^* \| \geqslant T',$

$$
E_{f_j^*}(\|t_i(X) - f_i^*\|^2) \leq c_j \|m_i^* - m_j^*\|^{p_j}, \quad j \neq i,
$$

\n
$$
C_i, \quad p_i
$$
\n(50)

B E . (47) (50),
\n
$$
E(||t_i(X) - f_i^*||^2) = \sum_{j=1}^K a_j^* E_{f_j^*} (||t_i(X) - f_i^*||^2) \le u M_i^q(F^*),
$$
\n
$$
M_i(F^*) = \sum_{j \ne i}^K ||m_i^* - m_j^*||, u \ne 0
$$

Proof of Lemma 3.

$$
f(Z) = o(Zp),
$$

\n
$$
Z \rightarrow 0, \qquad p
$$

\n
$$
F^*
$$

\n
$$
m_{ij}^*
$$

\n
$$
a_i^* P_i(m_{ij}^* | f_i^*) = a_j^* P_j(m_{ij}^* | f_j^*).
$$

\n
$$
Z(F^*) = Z. \qquad i \neq j,
$$

\n
$$
m_i^* \qquad m_j^*
$$

\n
$$
Z(F^*) = Z. \qquad i \neq j,
$$

\n
$$
Z(F^*) = Z. \qquad i \neq j,
$$

\n
$$
Z(F^*) = Z. \qquad i \neq j,
$$

$$
E_{i} = \{x : a_{i}^{*}P_{i}(x|f_{i}^{*}) \ge a_{j}^{*}P_{j}(x|f_{j}^{*})\},
$$
\n
$$
E_{j} = \{x : a_{j}^{*}P_{j}(x|f_{j}^{*}) > a_{i}^{*}P_{i}(x|f_{i}^{*})\}.
$$
\nA $Z(F^{*})$ \n
$$
Z
$$
\n<

$$
y_i, r_i \t r_j \t ||m_i^* - m_j^*|| \t ||m_i^* - m_j^*||
$$

 $b₂$

$$
r_i \geq b_i ||m_i^* - m_j^*||
$$
 $r_j \geq b_j ||m_i^* - m_j^*||.$

$$
\mathcal{D}_i = \mathcal{N}_{r_i}^c(m_i^*) = \{x \colon ||x - m_i^*|| \ge r_i\},
$$

$$
\mathcal{D}_j = \mathcal{N}_{r_j}^c(m_j^*) = \{x \colon ||x - m_j^*|| \ge r_j\}
$$

$$
E_i \subset D_j, \qquad E_j \subset D_i.
$$

\nM $\qquad , \qquad e_{ij}(\mathbf{F}^*) \qquad h_k(x)$
\n
$$
e_{ij}(\mathbf{F}^*) = \int h_i(x)h_j(x)P(x|\mathbf{F}^*) \mathbf{m}
$$

\n
$$
= \int_{E_i} h_i(x)h_j(x)P(x|\mathbf{F}^*) \mathbf{m} + \int_{E_j} h_i(x)h_j(x)P(x|\mathbf{F}^*) \mathbf{m}
$$

\n
$$
\leq \int_{\mathcal{D}_j} h_i(x)h_j(x)P(x|\mathbf{F}^*) \mathbf{m} + \int_{\mathcal{D}_i} h_i(x)h_j(x)P(x|\mathbf{F}^*) \mathbf{m}
$$

\n
$$
\leq \int_{\mathcal{D}_j} h_j(x)P(x|\mathbf{F}^*) \mathbf{m} + \int_{\mathcal{D}_i} h_i(x)P(x|\mathbf{F}^*) \mathbf{m}
$$

\n
$$
= \mathbf{a}_j^* \int_{\mathcal{D}_j} P_j(x|\mathbf{f}_j^*) \mathbf{m} + \mathbf{a}_i^* \int_{\mathcal{D}_i} P_i(x|\mathbf{f}_i^*) \mathbf{m}
$$

\n
$$
\int_{\mathcal{D}_i} P_i(x|\mathbf{f}_i^*) \mathbf{m} = \int_{\|x - m_i^* \| \leq b_i \|m_i^* - m_j^* \|}{\int_{\mathcal{D}_i} P_i(x|\mathbf{f}_i^*) \mathbf{m}}
$$

\nB $\qquad y = (x - m_i^*)/ \|m_i^* - m_j^* \|,$
\n
$$
\int_{\mathcal{D}_i} P_i(x|\mathbf{f}_i^*) \mathbf{m}
$$

\n
$$
\leq \int_{\|y\| \leq b_i} w(\|m_i^* - m_j^* \|y + m_i^*)(1^i) e^{-(1 - \frac{r(\|m_i^* - m_j^* \|)^2}{2})/(\frac{1}{\|y\|^2})} \frac{m_i^* - m_j^* \|}{m_i^* - m_j^* \|y - m_j^* \|}.
$$

\n
$$
= \int_{\|y\| \leq b_i} \|m_i^* - m_j^* \|y + m_i^*)(
$$

$$
\overline{m}
$$

$$
m_i^* = m_j^* \|y + m_i^*\)
$$

\n
$$
\|m_i^* - m_j^*\|^{-q} w(\|m_i^* - m_j^*\|y + m_i^*)
$$

\n
$$
Z(F^*) \to 0.
$$

$$
||m_i^* - m_j^*||^{-q} w(||m_i^* - m_j^*||y + m_i^*)
$$

\n
$$
||m_i^* - m_j^*||^{-q} (1^{i})^{-c_1} \le O(Z^{-c_1}),
$$

\n
$$
||m_i^* - m_j^*||^{-2} (1^{i})^{-nc_2} \ge O(Z^{-c_2}),
$$

\n
$$
c'_1 = (q+1) \vee (c_1/n).
$$

\nA
\n
$$
\int_{\mathcal{D}_i} P_t(x | f_i^*) \text{ } m \le \int_{\mathcal{B}_i} \frac{1}{Z^c_1(F^*)} w_1(y) - r'(1/Z^c_2(F^*))|y||^{c_2} \text{ } m'
$$

\n
$$
= \int_{\mathcal{B}_i} \frac{1}{Z^c_1} w_1(y) - r'(1/Z^c_2)|y||^{c_2} \text{ } m',
$$

\n
$$
\mathcal{B}_i = \{y : ||y|| \ge b_i\}, r'
$$

\nF
\n
$$
F,
$$

\n
$$
F_i(Z) = \int_{\mathcal{B}_i} P(y|Z) y, \quad P(y|Z) = \frac{1}{Z^{c_1}} w_1(y) - r'(1/Z^c_2)|y||^{c_2}
$$

\nF
\n
$$
y \in \mathcal{B}_i,
$$

\n
$$
\frac{P(y|Z)}{Z'} = w_1(y) \frac{1}{Z_{\rightarrow 0} \frac{C^{c'_1 + p}}{Z^{c'_1 + p}}} - r'(1/Z^c_2)|y||^{c_2}
$$

\n
$$
= 0,
$$

\n
$$
\mathcal{B}_i,
$$

\n
$$
\frac{F_2(Z)}{Z} = \frac{P(y|Z)}{Z} \text{ } m'
$$

\n
$$
= \int_{\mathcal{B}_i} \frac{P(y|Z)}{Z} \text{ } m'
$$

\n
$$
= \int_{\mathcal{B}_i} \frac{P(y|Z)}{Z} \text{ } m'
$$

\n
$$
= \int_{\mathcal{B}_i} \frac{P(y|Z)}{Z} \text{ } m'
$$

\n
$$
= 0
$$

\n
$$
F_i(Z
$$

 $A \qquad ,$

$$
f_{ij}(\mathbf{Z}) = \sum_{\mathbf{Z}(\mathbf{F}^*)=Z} e_{ij}(\mathbf{F}^*)
$$

\n
$$
\leq \sum_{\mathbf{Z}(\mathbf{F}^*)=Z} \left(\mathbf{a}_j^* \int_{\mathcal{D}_j} P_j(x|\mathbf{f}_j^*) \mathbf{m} + \mathbf{a}_i^* \int_{\mathcal{D}_i} P_i(x|\mathbf{f}_i^*) \mathbf{m} \right)
$$

\n
$$
\leq \sum_{\mathbf{Z}(\mathbf{F}^*)=Z} \int_{\mathcal{D}_j} P_j(x|\mathbf{f}_j^*) \mathbf{x} + \sum_{\mathbf{Z}(\mathbf{F}^*)=Z} \int_{\mathcal{D}_i} P_i(x|\mathbf{f}_i^*) \mathbf{m}
$$

\n
$$
= o(\mathbf{Z}^p).
$$

\n,
\n
$$
f(\mathbf{Z}) \leq \int_{\mathcal{U}^*} f_{ij}(\mathbf{Z}) = o(\mathbf{Z}^p).
$$

\n
$$
M
$$

 $z\rightarrow 0$ $\frac{f^{e}(Z)}{Z^{p}} = \sum_{Z \to 0}$ $f(\mathbf{Z})$ $\mathsf{Z}_{\mathsf{e}}^{\scriptscriptstyle\mathsf{p}}$ $\angle f(T)$ ^e $= 0,$ $f^{\mathbf{e}}(Z) = o(Z^p)$ $f^{\mathbf{e}}(Z(F^*)) = o(Z^p(F^*))$.

References

information theory.

Lei Xu (IEEE Fellow) is a chair professor of $\& E$, and $\& E$, and $\& E$, and $\& E$, and $\& E$ Chinese University of Hong Kong. He completed his Ph.D. thesis at Tsinghuan Ph.D. the Tsinghuan Ph.D. the Tsinghuan Ph 1986, the end of 1986 , then \overline{D} Mathematics, Peking of Mathematics, Peking o 1987 $\begin{array}{ccc} \text{P6} & \text{P6} & \text{P6} \\ \text{F} & \text{P6} & \text{A}, & \text{H} \\ \text{P6} & \text{P6} & \text{P6} \end{array}$ MI . H C HK

 F , Canada and MIT. He joint MIT. He joint MIT. He joint matrix H including Harvard and MIT. He joint matrix H including Harvard and MIT. He joint matrix H is a contract of the including Harvard and MIT. He joint m $\begin{array}{ccc} &\textrm{A},&\textrm{H}\\ &\ddots&\\ \end{array}$ 2002. P 2002. Prof. \blacksquare

, including Neural Networks, IEEE Transactions on Neural Neural Networks, as \sim Neural Neural Neural Networks, as \sim $\mathbf{N} = \mathbf{N}$ $\mathbf{N} = (01-03),$ C Finance Telecomputational Committee of I (01-03), Γ

A -P N N A \mathbf{A} P i 100 , pattern recognition, neurocomputing, pattern recognition, neurocomputing, and computing, and computing, and computation, η , with a number of the fields. He has given a number of \mathcal{H}_1

 $(1994 \, \text{C} \quad N \quad N \quad A)$ (including 1995 INN Leadership August 2012 IEEE Fellow and a Fellow and a Terminalism and a Fellow of In A and P , and A and A member of E and A