Modelos de misturas de gaussianas (GMMs)

Miguel Arjona Ramírez maramire@usp.br

Escola Politécnica

Universidade de São Paulo

Roteiro

1. Componentes da mistura e densidade composta.
2. Introdução de uma variável latente vetorial discreta.
3. Densidades condicional, conjunta e marginal reconstruída.
4. Função de log verossimilhança e possíveis singularidades.
5. Estimação de GMM pelo algoritmo da Esperança-Maximização (EM)

Modelos de misturas de gaussianas

1.
$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k N\left(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$$

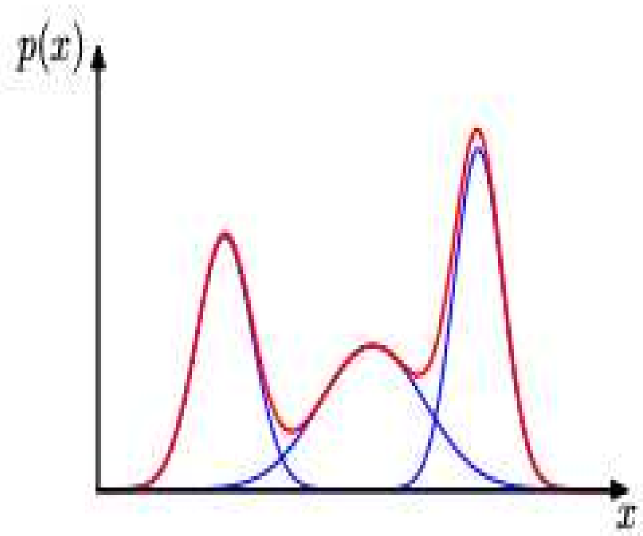
2. $\boldsymbol{z}$: VA binária K-dimensional 1-de-K com $z_k \in \{0,1\}$ e $\sum_{k=1}^{K} z_k = 1$.

3. Função massa de probabilidade (fmp) da VA $\boldsymbol{z}$ é dada pelos coeficientes da mistura

$$P\{z_k = 1\} = \pi_k$$

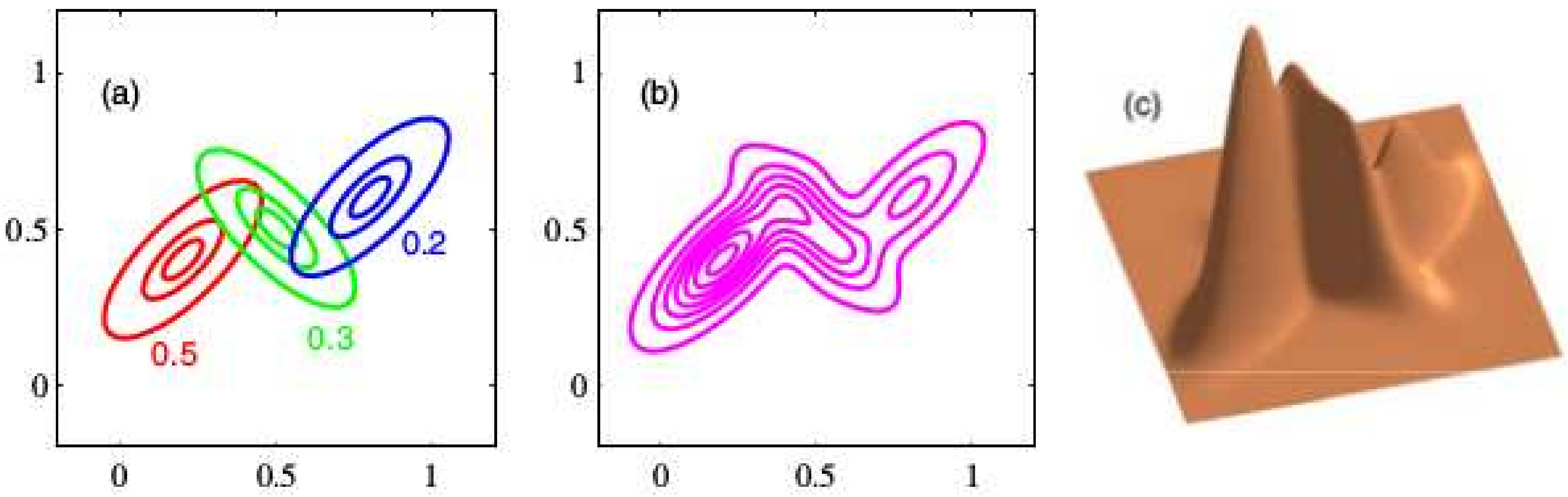
com as $K+1$ condições $0 \leq \pi_k \leq 1$ e $\sum_{k=1}^{K} \pi_k = 1$.

GMM com as componentes gaussianas



PATTERN RECOGNITION and MACHINE LEARNING, Christofer M. Bishop, 2006, Springer

GMM como uma densidade em si



Massa de probabilidade da variável latente

1. Função massa de probabilidade (fmp) da VA $\boldsymbol{z}$ é dada pelos coeficientes da mistura

$$P\{z_k = 1\} = \pi_k$$

com as $K+1$ condições $0 \leq \pi_k \leq 1$ e $\sum_{k=1}^{K} \pi_k = 1$.

2.

$$p(\boldsymbol{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$$

Densidade condicional

1.

$$p(\boldsymbol{x}|z_k = 1) = N\left(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$$

2.

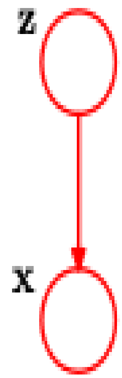
$$p(\boldsymbol{x}|\boldsymbol{z}) = \prod_{k=1}^{K} N^{z_k}\left(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$$

Densidades conjunta e marginal

$$
\begin{aligned}
p(\boldsymbol{x}, \boldsymbol{z}) &= p(\boldsymbol{z}) p(\boldsymbol{x}|\boldsymbol{z}) \\
&= \prod_{k=1}^{K} \pi_k^{z_k} \prod_{k=1}^{K} N^{z_k} \left(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
p(\boldsymbol{x}) &= \sum_{\boldsymbol{z}} p(\boldsymbol{x}, \boldsymbol{z}) \\
&= \sum_{k=1}^{K} \pi_k N \left(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)
\end{aligned} \tag{2}
$$

GMM com variável latente



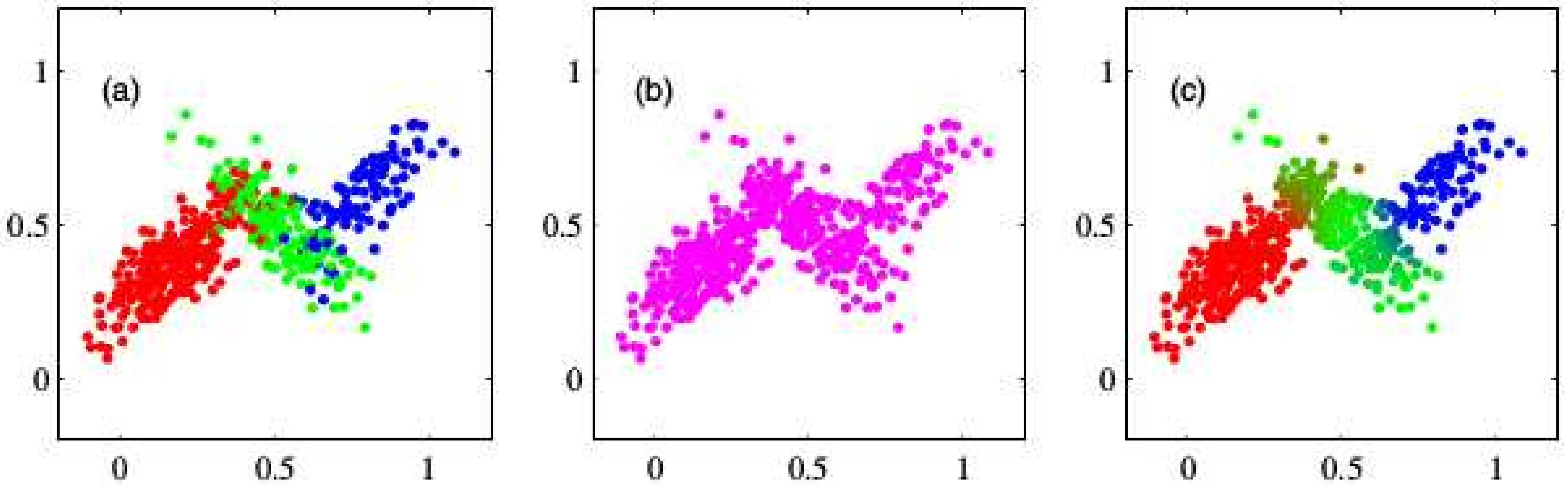
PATTERN RECOGNITION and MACHINE LEARNING, Christofer M. Bishop, 2006, Springer

Densidades a posteriori

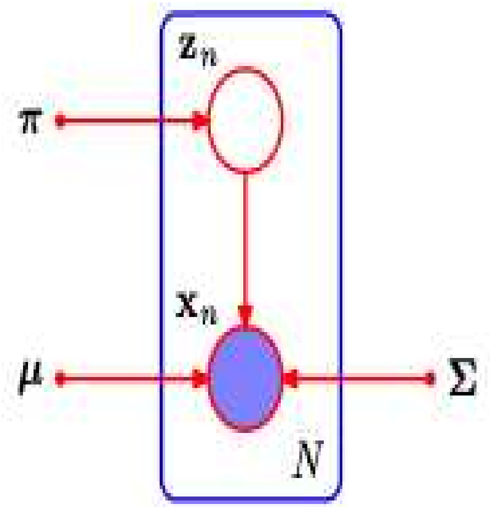
Pelo Teorema de Bayes, obtemos

$$\gamma(z_k) \equiv p(z_k = 1|\boldsymbol{x}) = \frac{p(z_k = 1)p(\boldsymbol{x}|z_k = 1)}{\sum_{i=1}^{K} p(z_i = 1)p(\boldsymbol{x}|z_i = 1)} = \frac{\pi_k N\left(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)}{\sum_{i=1}^{K} \pi_i N\left(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)} \quad (3)$$

GMM de $K = 3$ componentes e $N = 500$ dados



Grafo de GMM com N dados para treino

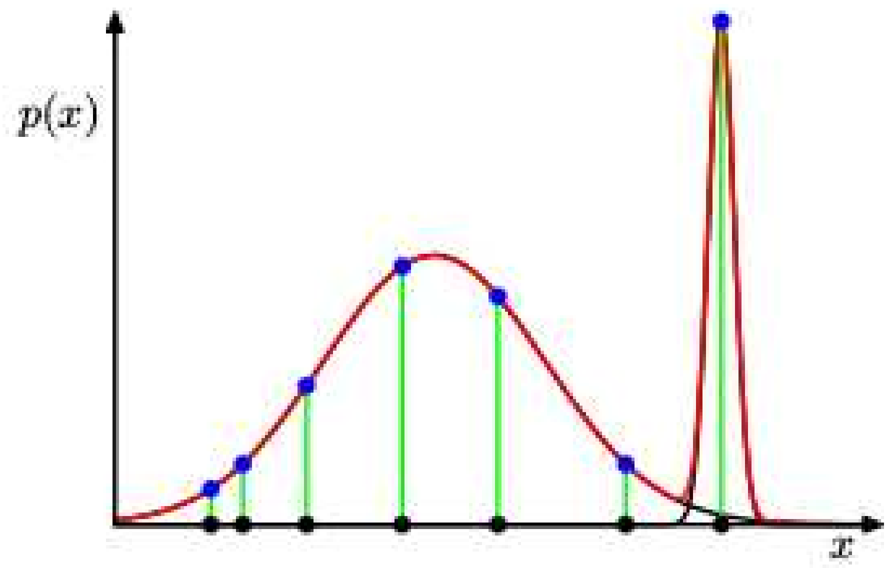


PATTERN RECOGNITION and MACHINE LEARNING, Christofer M. Bishop, 2006, Springer

Função de verossimilhança de GMM

$$\ln p(\boldsymbol{x}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \ln \left[\sum_{k=1}^{K} \pi_k N(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right]$$

Singularidades na verossimilhança de GMM



Se $\boldsymbol{\mu}_j = \boldsymbol{x}_n$, com $\boldsymbol{\Sigma}_j = \sigma_j^2 \boldsymbol{I}$, a verossimilhança tem uma parcela

$$N\left(\boldsymbol{x}_n | \boldsymbol{x}_n, \sigma_j^2 \boldsymbol{I}\right) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{\sigma_j^D}$$

PATTERN RECOGNITION and MACHINE LEARNING, Christofer M. Bishop, 2006, Springer

Esperança-Maximização (EM) para GMM

$L = \sum_{n=1}^{N} \ln G$: log verossimilhança

G : GMM

$$N_k = \frac{1}{(2\pi)^{\frac{D}{2}} \det^{\frac{1}{2}}(\boldsymbol{\Sigma}_k)} \exp\left[-\frac{1}{2}(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu}_k)\right]$$

Gaussiana

A_k : Expoente de gaussiana

$$\boldsymbol{\nabla}_{\boldsymbol{\mu}_k} L \equiv \frac{\partial L}{\partial \boldsymbol{\mu}_k} = \frac{\partial L}{\partial G}\frac{\partial G}{\partial N_k}\frac{\partial N_k}{\partial A_k}\frac{\partial A_k}{\partial \boldsymbol{\mu}_k}$$

$$\frac{\partial L}{\partial G} = \sum_{n=1}^{N}\frac{1}{G}, \quad \frac{\partial G}{\partial N_k} = \pi_k, \quad \frac{\partial N_k}{\partial A_k} = N_k, \quad \frac{\partial A_k}{\partial \boldsymbol{\mu}_k} = \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu}_k).$$

Maximização (M) para as médias do GMM (1)

$$\nabla_{\boldsymbol{\mu}_k} L = \frac{\partial L}{\partial G} \frac{\partial G}{\partial N_k} \frac{\partial N_k}{\partial A_k} \frac{\partial A_k}{\partial \boldsymbol{\mu}_k}$$

$$\nabla_{\boldsymbol{\mu_k}} L = \sum_{n=1}^{N} \frac{1}{G} \pi_k N_k \boldsymbol{\Sigma_k}^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu}_k) = \mathbf{0}$$

$$\sum_{n=1}^{N} \frac{\pi_k N(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma_k})}{\sum_{i=1}^{K} N(\boldsymbol{x}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma_i})} \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu}_k) = \mathbf{0}$$

$$\sum_{n=1}^{N} \gamma(z_{nk}) \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu}_k) = \mathbf{0}$$

Maximização (M) para as médias do GMM (2)

$$\sum_{n=1}^{N} \gamma(z_{nk}) \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu}_k) = \mathbf{0}$$

$$\boldsymbol{\mu}_k \sum_{n=1}^{N} \gamma(z_{nk}) = \sum_{n=1}^{N} \gamma(z_{nk}) \boldsymbol{x}_n$$

$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) \boldsymbol{x}_n}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) \boldsymbol{x}_n}{N_k}$$

Etapa M para as covariâncias do GMM (1)

$$\nabla_{\boldsymbol{\Sigma}_k} L = \frac{\partial L}{\partial G}\frac{\partial G}{\partial N_k}\left[\frac{\partial N_k}{\partial F_k}\frac{\partial F_k}{\partial D_k}\frac{\partial D_k}{\partial \boldsymbol{\Sigma}_k} + \frac{\partial N_k}{\partial A_k}\frac{\partial A_k}{\partial \boldsymbol{\Sigma}_k}\right]$$

$$F_k = \frac{1}{D_k}, \quad D_k = (2\pi)^{\frac{D}{2}} \det{}^{\frac{1}{2}}(\boldsymbol{\Sigma}_k).$$

$$\frac{\partial N_k}{\partial F_k} = \exp(A_k), \quad \frac{\partial F_k}{\partial D_k} = -\frac{1}{D_k^2}.$$

Etapa M para as covariâncias do GMM (2)

$$\nabla_{\boldsymbol{\Sigma}_k} L = \frac{\partial L}{\partial G}\frac{\partial G}{\partial N_k}\left[\frac{\partial N_k}{\partial F_k}\frac{\partial F_k}{\partial D_k}\frac{\partial D_k}{\partial \boldsymbol{\Sigma}_k} + \frac{\partial N_k}{\partial A_k}\frac{\partial A_k}{\partial \boldsymbol{\Sigma}_k}\right]$$

$$\begin{aligned}
\frac{\partial D_k}{\partial \boldsymbol{\Sigma}_k} &= (2\pi)^{\frac{D}{2}}\frac{1}{2}\det{}^{-\frac{1}{2}}(\boldsymbol{\Sigma}_k)\frac{\partial \det(\boldsymbol{\Sigma}_k)}{\partial \boldsymbol{\Sigma}_k} \\
&= (2\pi)^{\frac{D}{2}}\frac{1}{2}\det{}^{-\frac{1}{2}}(\boldsymbol{\Sigma}_k)\det(\boldsymbol{\Sigma}_k)\boldsymbol{\Sigma}_k^{-1} \\
&= \frac{1}{2}(2\pi)^{\frac{D}{2}}\det{}^{\frac{1}{2}}(\boldsymbol{\Sigma}_k)\boldsymbol{\Sigma}_k^{-1} \\
&= \frac{1}{2}D_k\boldsymbol{\Sigma}_k^{-1}.
\end{aligned}$$

Derivada matricial do próprio determinante

$$\boldsymbol{B} = \frac{\partial \det(\boldsymbol{A})}{\partial \boldsymbol{A}}$$

$$\det(\boldsymbol{A}) = \sum_{j=1}^{D} (-1)^{i+j} a_{ij} M_{ij} = \sum_{i=1}^{D} (-1)^{i+j} a_{ij} M_{ij}$$

$$\boldsymbol{B} = \mathsf{adj}^T(\boldsymbol{A})$$

$$\boldsymbol{A}^{-1} = \det{}^{-1}(\boldsymbol{A})\mathsf{adj}(\boldsymbol{A})$$

$$\boldsymbol{B} = \det(\boldsymbol{A})\boldsymbol{A}^{-T}.$$

Etapa M para as covariâncias do GMM (3)

$$\nabla_{\boldsymbol{\Sigma}_k} L = \frac{\partial L}{\partial G} \frac{\partial G}{\partial N_k} \left[\frac{\partial N_k}{\partial F_k} \frac{\partial F_k}{\partial D_k} \frac{\partial D_k}{\partial \boldsymbol{\Sigma}_k} + \frac{\partial N_k}{\partial A_k} \frac{\partial A_k}{\partial \boldsymbol{\Sigma}_k} \right]$$

$$\begin{aligned} \frac{\partial A_k}{\partial \boldsymbol{\Sigma}_k} &= -\frac{1}{2} \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k \right) \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k \right)^T \frac{\partial \boldsymbol{\Sigma}_k^{-1}}{\partial \boldsymbol{\Sigma}_k} \\ &= \frac{1}{2} \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k \right) \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k \right)^T \left(\boldsymbol{\Sigma}_k^{-1} \right)^2 . \end{aligned}$$

$$\frac{\partial L}{\partial G} = \sum_{n=1}^{N} \frac{1}{G}, \quad \frac{\partial G}{\partial N_k} = \pi_k, \quad \frac{\partial N_k}{\partial A_k} = N_k, \quad \frac{\partial A_k}{\partial \boldsymbol{\mu}_k} = \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu}_k).$$

Derivada da matriz inversa

$$\mathbf{0} = \boldsymbol{I}' = \boldsymbol{A}' \cdot \boldsymbol{A}^{-1} + \boldsymbol{A} \cdot \left(\boldsymbol{A}^{-1}\right)'$$

$$\boldsymbol{A} \cdot \left(\boldsymbol{A}^{-1}\right)' = -\boldsymbol{A}' \cdot \boldsymbol{A}^{-1}$$

$$\left(\boldsymbol{A}^{-1}\right)' = -\boldsymbol{A}^{-1} \cdot \boldsymbol{A}' \cdot \boldsymbol{A}^{-1}$$

Etapa M para as covariâncias do GMM (4)

$$\nabla_{\boldsymbol{\Sigma}_k} L = \frac{\partial L}{\partial G} \frac{\partial G}{\partial N_k} \left[\frac{\partial N_k}{\partial F_k} \frac{\partial F_k}{\partial D_k} \frac{\partial D_k}{\partial \boldsymbol{\Sigma}_k} + \frac{\partial N_k}{\partial A_k} \frac{\partial A_k}{\partial \boldsymbol{\Sigma}_k} \right]$$

$$\nabla_{\boldsymbol{\Sigma}_k} L = \sum_{n=1}^{N} \frac{1}{G} \pi_k \left[D_k N_k \left(-\frac{1}{D_k^2} \right) \frac{1}{2} D_k \boldsymbol{\Sigma}_k^{-1} \right.$$
$$\left. + N_k \cdot \left(\frac{1}{2} \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k \right) \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k \right)^T \left(\boldsymbol{\Sigma}_k^{-1} \right)^2 \right) \right].$$

Etapa M para as covariâncias do GMM (5)

$$\nabla_{\boldsymbol{\Sigma}_k} L = \mathbf{0}$$

$$\sum_{n=1}^{N} \frac{\pi_k N_k}{G} \boldsymbol{\Sigma}_k = \sum_{n=1}^{N} \frac{\pi_k N_k}{G} \cdot \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k\right)\left(\boldsymbol{x}_n - \boldsymbol{\mu}_k\right)^T$$

$$\sum_{n=1}^{N} \gamma(z_{nk}) \boldsymbol{\Sigma}_k = \sum_{n=1}^{N} \gamma(z_{nk}) \cdot \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k\right)\left(\boldsymbol{x}_n - \boldsymbol{\mu}_k\right)^T .$$

$$\boldsymbol{\Sigma}_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) \cdot \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k\right)\left(\boldsymbol{x}_n - \boldsymbol{\mu}_k\right)^T}{N_k}$$

Etapa M para os coeficientes da mistura (1)

$$\nabla_{\pi_k} L = \frac{\partial L}{\partial G} \frac{\partial G}{\partial \pi_k}$$

Restrição: $\sum_{i=1}^{K} \pi_i - 1 = 0$

Lagrangiano: $J = L + \lambda \left(\sum_{i=1}^{K} \pi_i - 1 \right)$

$$\nabla_{\pi_k} J = \frac{\partial L}{\partial G} \frac{\partial G}{\partial \pi_k} + \lambda$$

$$\nabla_{\pi_k} L = 0$$

Etapa M para os coeficientes da mistura (2)

$$\nabla_{\pi_k} J = \frac{\partial L}{\partial G} \frac{\partial G}{\partial \pi_k} + \lambda$$

$$\nabla_{\pi_k} L = 0$$

$$\lambda = -\sum_{n=1}^{N} \frac{1}{G} N_k = -\frac{\sum_{n=1}^{N} \gamma(z_{nk})}{\pi_k}$$

$$\lambda \sum_{k=1}^{K} \pi_k = \sum_{n=1}^{N} \frac{\sum_{k=1}^{K} \pi_k N_k}{G}.$$

$$\lambda = -N.$$

Etapa M para os coeficientes da mistura (3)

$$\nabla_{\pi_k} J = \frac{\partial L}{\partial G} \frac{\partial G}{\partial \pi_k} + \lambda$$

$$\lambda = -\sum_{n=1}^{N} \frac{1}{G} N_k = -\frac{\sum_{n=1}^{N} \gamma(z_{nk})}{\pi_k}$$

$$\lambda = -N.$$

$$\pi_k = \frac{N_k}{N}.$$

Algoritmo EM para GMM (i), (ii)

(*i*)Inicialize as médias $\boldsymbol{\mu}_k$, as covariâncias Σ_k e os pesos das componentes π_k e calcule o valor inicial da log verossimilhança.

(*ii*)Etapa E: Calcule as responsabilidades ou probabilidades a posteriori usando os valores atuais dos parâmetros

$$\gamma(z_{nk}) = \frac{\pi_k N(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{i=1}^{K} \pi_i N(\boldsymbol{x}_n|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}$$

Algoritmo EM para GMM (iii)

(*iii*)Etapa M: Reestime os parâmetros usando as responsabilidades ou probabilidades atuais

$$\boldsymbol{\mu}_k^{\text{nova}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \boldsymbol{x}_n$$

$$\boldsymbol{\Sigma}_k^{\text{nova}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k^{\text{nova}}\right) \left(\boldsymbol{x}_n - \boldsymbol{\mu}_k^{\text{nova}}\right)^T$$

$$\pi_k^{\text{novo}} = \frac{N_k}{N},$$

sendo $N_k = \sum_{n=1}^{N} \gamma(z_{nk})$.

Algoritmo EM para GMM (iv)

(*iv*)Calcule a log verossimilhança

$$\ln p(\boldsymbol{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \sum_{n=1}^{N} \ln \left[\sum_{k=1}^{K} \pi_k N(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right]$$

e verifique a convergência dos parâmetros ou da log verossimilhança. Se não for satisfeito o critério de convergência, retorne à etapa *(ii).*