PGMAT0061/MATG17 - Técnicas Computacionais em Estatística "Simulação Estocástica"

Professor responsável: Paulo Henrique Ferreira da Silva

Universidade Federal da Bahia
Instituto de Matemática e Estatística
Departamento de Estatística

23 de Janeiro de 2025

Algoritmo EM

Motivação:

- O algoritmo EM (do inglês *Expectation-Maximization*) pode ser aplicado em modelos de dados faltantes (*missing data models*);
- Esses modelos têm uma função de verossimilhança que pode ser expressa por:

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = \int_{\mathcal{Z}} f(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}) d\boldsymbol{z}$$

Isto é, a função de interesse é a <u>marginal</u> de um modelo conjunto para $(\boldsymbol{X}, \boldsymbol{Z})$.

Algoritmo EM

- É um método iterativo que calcula o estimador de máxima verossimilhança (EMV) em casos de dados faltantes ou incompletos;
- Transforma o problema de encontrar o máximo de uma função complicada em 2 problemas mais simples: esperança (E) e maximização (M);
- Foi introduzido por Dempster, Laird e Rubin (1977).

Algoritmo EM

Seja $\boldsymbol{X} = (X_1, \ldots, X_n)$ uma a.a. de $X \sim f(x|\boldsymbol{\theta})$, com $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$.

A função de verossimilhança é definida por:

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})$$

Queremos encontrar $\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}|\boldsymbol{x})$.

Algoritmo EM

Considere as variáveis auxiliares (ou dados faltantes) $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ e seja $f(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta})$ a distribuição conjunta para $(\boldsymbol{X}, \boldsymbol{Z})$.

- Verossimilhança para dados completos:

$$L^c(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{z}) = f(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta})$$

- Verossimilhança para dados incompletos:

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = f(\boldsymbol{x}|\boldsymbol{\theta}) = \int_{\mathcal{Z}} f(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}) d\boldsymbol{z}$$

Sabemos que: $f(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{\theta}) = \dfrac{f(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta})}{f(\boldsymbol{x}|\boldsymbol{\theta})}$. Logo,

$$\begin{aligned} \log f(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{\theta}) &= \log f(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}) - \log f(\boldsymbol{x}|\boldsymbol{\theta}) \\ &= \ell^c(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{z}) - \ell(\boldsymbol{\theta}|\boldsymbol{x}) \end{aligned}$$

Algoritmo EM

Tomando a esperança na distribuição de $\boldsymbol{Z}|\boldsymbol{x}, \boldsymbol{\theta}$ para $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, obtemos:

$$\ell(\boldsymbol{\theta}|\boldsymbol{x}) = \mathbb{E}[\ell^c(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{z})|\boldsymbol{x}, \boldsymbol{\theta}_0] - \mathbb{E}[\log f(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{\theta})|\boldsymbol{x}, \boldsymbol{\theta}_0]$$

Maximizar $\ell(\boldsymbol{\theta}|\boldsymbol{x})$ equivale a maximizar somente:

$$\mathbb{E}[\ell^c(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{z})|\boldsymbol{x}, \boldsymbol{\theta}_0]$$

Algoritmo EM

Algoritmo EM:

1. (Passo E) Calcular: $Q(\boldsymbol{\theta}|\boldsymbol{x}, \widehat{\boldsymbol{\theta}}_{(k)}) = \mathbb{E}[\ell^c(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{z})|\boldsymbol{x}, \widehat{\boldsymbol{\theta}}_{(k)}]$
2. (Passo M) Calcular: $\widehat{\boldsymbol{\theta}}_{(k+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{x}, \widehat{\boldsymbol{\theta}}_{(k)})$

<u>OBS:</u> Repetir (1)-(2) até obter convergência, por exemplo, $||\widehat{\boldsymbol{\theta}}_{(k+1)} - \widehat{\boldsymbol{\theta}}_{(k)}|| < \epsilon$, em que ϵ é um valor determinado maior que 0.

Algoritmo EM

Podemos mostrar que:

$$L(\widehat{\boldsymbol{\theta}}_{(k+1)}|\mathbf{x}) \geq L(\widehat{\boldsymbol{\theta}}_{(k)}|\mathbf{x})$$

com igualdade se, e somente se,

$$Q(\widehat{\boldsymbol{\theta}}_{(k+1)}|\mathbf{x}, \widehat{\boldsymbol{\theta}}_{(k)}) = Q(\widehat{\boldsymbol{\theta}}_{(k)}|\mathbf{x}, \widehat{\boldsymbol{\theta}}_{(k)})$$

Algoritmo EM

Exemplo: (modelo de mistura binomial)

Considere duas moedas com probabilidades de cara (desconhecidas), denotadas por p e q, respectivamente. A 1ª moeda é escolhida com probabilidade π e a 2ª moeda é escolhida com probabilidade $1 - \pi$. A moeda escolhida é lançada e o resultado é registrado.

Ex: $\boldsymbol{x} = \{1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1\}$, com: $1 =$ "cara", $0 =$ "coroa"

Seja $Z_i \in \{0, 1\}$ a moeda que foi usada no i-ésimo lançamento, $i = 1, \ldots, n$ ($0 =$ "moeda 1", $1 =$ "moeda 2")

Solução:

$$\boldsymbol{\theta} = (p, q, \pi)$$

$$\widehat{\boldsymbol{\theta}}_{\text{MV}} = \arg\max_{\boldsymbol{\theta}} \log p(\boldsymbol{x}|\boldsymbol{\theta})$$

Algoritmo EM

1. **Passo E:**

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= \mathbb{E}\left[\log \prod_{i=1}^{n} \left\{\pi p^{x_i}(1-p)^{1-x_i}\right\}^{z_i} \left\{(1-\pi)q^{x_i}(1-q)^{1-x_i}\right\}^{1-z_i}\right] \\
&= \sum_{i=1}^{n} \left\{\mathbb{E}\left[Z_i|x_i,\boldsymbol{\theta}^{(k)}\right]\left(\log\pi + x_i\log p + (1-x_i)\log(1-p)\right)\right. \\
&\quad \left. + \left(1-\mathbb{E}\left[Z_i|x_i,\boldsymbol{\theta}^{(k)}\right]\right)\left(\log(1-\pi) + x_i\log q + (1-x_i)\log(1-q)\right)\right\}
\end{aligned}
$$

$$
\begin{aligned}
\mu_i^{(k)} = \mathbb{E}\left[Z_i|x_i,\boldsymbol{\theta}^{(k)}\right] &= p(z_i = 1|x_i,\boldsymbol{\theta}^{(k)}) \\
&= \frac{p(x_i|z_i=1,\boldsymbol{\theta}^{(k)})p(z_i=1|\boldsymbol{\theta}^{(k)})}{p(x_i|\boldsymbol{\theta}^{(k)})} \\
&= \frac{\pi^{(k)}[p^{(k)}]^{x_i}[1-p^{(k)}]^{1-x_i}}{\pi^{(k)}[p^{(k)}]^{x_i}[1-p^{(k)}]^{1-x_i} + (1-\pi^{(k)})[q^{(k)}]^{x_i}[1-q^{(k)}]^{1-x_i}}
\end{aligned}
$$

Algoritmo EM

2. **Passo M:** (exercício)

$$\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})}{\partial \pi} = 0 \quad \Longrightarrow \quad \pi^{(k+1)} = \frac{1}{n}\sum_i \mu_i^{(k)}$$

$$\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})}{\partial p} = 0 \quad \Longrightarrow \quad p^{(k+1)} = \frac{\sum_i \mu_i^{(k)} x_i}{\sum_i \mu_i^{(k)}}$$

$$\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})}{\partial q} = 0 \quad \Longrightarrow \quad q^{(k+1)} = \frac{\sum_i (1-\mu_i^{(k)}) x_i}{\sum_i (1-\mu_i^{(k)})}$$