Trabalho II: o algoritmo EM.

Disciplina: Inferência Estatística
Aluno: Iara Cristina Mescua Castro
Curso: Matemática Aplicada

27 de setembro de 2022

Data de Entrega: 28 de Setembro de 2022.

Questões

1. Defina claramente todos os passos do algoritmo EM (faça um glossário de termos se precisar);

2. **Um exemplo motivador:** Suponha que temos duas moedas, Moeda 1 e Moeda 2 de modo que Pr(Cara | Moeda = 1) = p_1 e Pr(Cara | Moeda = 2) = p_2; Suponha agora que fazemos o seguinte experimento:

   (i) Selecionamos uma moeda aleatóriamente com probabilidade 1/2;
   (ii) Lançamos a moeda selecionada m vezes;
   (iii) Repetimos (i) e (ii) n vezes.

   Podemos representar os dados advindos deste experimento como

$$\begin{array}{ccccc} X_{11} & \dots & X_{1m} & & M_1 \\ X_{21} & \dots & X_{2m} & & M_2 \\ \vdots & \dots & \vdots & & \vdots \\ X_{n1} & \dots & X_{nm} & & M_n \end{array}.$$

   onde os X_{ij} são variáveis Bernoulli que guardam o resultados do lançamento da moeda e $M_i \in \{1, 2\}$ é a variável aleatória que guarda qual moeda foi selecionada na i-ésima rodada do experimento.

   Desenvolva um esquema EM para obter o EMV para $\theta = (p_1, p_2)$ quando desconhecemos os valores de M_i, isto é, quando não sabemos que moeda foi escolhida em cada uma das n rodadas.

3. Mostre que a sequência $\theta^{(j)}$ é monotônica e não descrescente com respeito à verossimilhança, isto é,

$$L\left(\theta^{(j+1)}\right) \geq L\left(\theta^{(j)}\right)$$

   Dica (fortemente aconselhado): ver exercício 7.31 de Casella & Berger (2002).

4. Discuta a importância do método EM: quando ele é aplicável? Vale sempre a pena? O que o item anterior demonstra sobre o método?

Respostas

1. Do ponto de vista frequentista, pode-se pensar que os dados observados são gerados a partir de variáveis aleatórias X conjuntamente com dados faltantes ou não observados Z.
   $Y = (X, Z)$ representa o vetor de dados completos.

   Dado os dados observados, x, deseja-se maximizar a função de verossimilhança $L(\theta, x)$. Frequentemente isto não é simples no entanto pode-se simplificar usando as densidades $Y|\theta$ e $Z|x, \theta$. O algoritmo EM faz isso.

   Seja $f_x(x|\theta)$ e $f_y(y|\theta)$ as distribuições dos dados observados e os dados completos, respectivamente

   $f_x(x|\theta) = \int f(y|\theta)dz$ e $f_{z|x}(z|x, \theta) = \frac{f_y(y|\theta)}{f_x(x|\theta)}$... (A)

   $$L(Y|\theta) = L(X, Z|\theta)$$

   O objetivo do algoritmo EM e encontrar o máximo de $L(Y|\theta)$ com relação a θ. Seja $\theta^{(t)}$ o máximo obtido na iteração t. t = 0,1,2,... Define-se por $Q(\theta, \theta^{(0)})$

   $$Q(\theta, \theta^{(t))} = E\{logL(\theta, Y)\} \;\; (1)$$
   $$= E\{\log f_y(y|\theta)|x, \theta^{(t)}\} \;\; (2)$$
   $$= \int [\log f_y(y|\theta)] \log f_{z|x}(z|x, \theta^{(t)})dz \;\; (3)$$

   Onde (3) estabelece que z é a única parte aleatória dado X = x.

   O algoritmo EM é inicializado em $O^{(0)}$ e alterna nos seguites passos:

   1) E STEP: Calcular $Q(\theta, \theta^{(0)})$.
   2) M STEP: Maximizar o $Q(\theta, \theta^{(0)})$.
   3) Retornar a 1 até convergência.

2. Seja x_j o resultado da j-ésima sequência de tamanho "n"

   $$P(x_i) = \frac{1}{2}P(x_i|M_1) + \frac{1}{2}P(x_i|M_2), \;\; j = 1, ...$$

   Dada a moeda $M_j \rightarrow$ cada elemento da sequência segue uma distribuição de Bernoulli com probabilidade p_1 ou p_2.

   - O problema é encontrar o EMV de $\theta = (p_1, p_2)$.
   - Considere $Y = (Z, \underline{X})$, onde:

   $$Y_1 = (Z_1, \underline{X}_1), Y_2 = (Z_2, \underline{X}_2), ..., Y_n = (Z_n, \underline{X}_n)$$

   Z_j é um indicador que identifica de que a moeda é M_1 ou M_2.

$$\rightarrow Z_j = 0 \Rightarrow M_1 \; ou \; Z_j = 1 \Rightarrow M_2$$

Logo $Z = (z_1, ..., z_n)$ corresponde a variável de alocação não observada. Assim a verossimilhança completa pode ser representada por:

$$\log L(\theta, Y) = \sum_{j=1}^{n} [(1 - z_j) \log P(\underline{\mathrm{X}}_i|M_1) + z_j \log P(\underline{\mathrm{X}}_i|M_2)]$$

Logo:

$$Q(\theta, \theta^{(t)}) = E\{\sum_{j=1}^{n} [(1 - z_j) \log P(\underline{\mathrm{X}}_i|M_1) + z_j \log P(\underline{\mathrm{X}}_i|M_2)]|\underset{\sim}{x}, \theta^{(t)}\}$$

$= E\{\sum_{j=1}^{n}(1-z_j)[\sum_{k=1}^{m} x_{jk} \log p_1 + m - \sum_{k=1}^{m} x_{jk} \log(1-p_1)] + \sum_{j=1}^{n} z_j [\sum_{k=1}^{m} x_{jk} \log p_2 + m - \sum_{k=1}^{m} x_{jk} \log(1 - p_2)]|\underset{\sim}{x}, \theta^{(t)}\}$... (α)

E-Step

Logo: Maximizar $Q(\theta, \theta^{(t)})$ equivale a maximizar $(\alpha)'$ com relacão a $\theta = (p_1, p_2)$ temos que o valor que maximiza α. E dado por:

$$P_1^{(t+1)} = \frac{\sum_{j=1}^{n}(1 - E(Z_j|\underset{\sim}{x}, \theta^{(t)}) \sum_{k=1}^{m} x_{jk}}{m \sum_{j=1}^{n})1 - E(Z_j|\underset{\sim}{x}, \theta^{(t)})}, \quad (*)$$

$$P_2^{(t+1)} = \frac{\sum_{j=1}^{n} E(Z_j|\underset{\sim}{x}, \theta^{(t)} \sum_{k=1}^{m} x_{jk}}{m \sum_{j=1}^{n} E(Z_j|\underset{\sim}{x}, \theta^{(t)})}, \quad (**)$$

Onde:

$$E(Z_j|\underline{\mathrm{x}}, \theta^{(t)}) = \frac{\frac{1}{2} P(\underset{\sim}{x_j}|M_1)}{\frac{1}{2} P(\underset{\sim}{x_j}|M_1) + \frac{1}{2} P(\underset{\sim}{x_j}|M_2)} \quad (* * *)$$

$$z_j^{(t+1)} = E(z_j|\underset{\sim}{x}, \theta^{(t)}) = \frac{\hat{p_2}^{\sum_{k=1}^{m} x_{jk}} (1 - \hat{p_2})^{m - \sum_{k=1}^{m} x_{jk}}}{\hat{p_1}^{\sum_{k=1}^{m} x_{jk}} (1 - \hat{p_1})^{m - \sum_{k=1}^{m} x_{jk}} + \hat{p_2}^{\sum_{k=1}^{m} x_{jk}} (1 - \hat{p_2})^{m - \sum_{k=1}^{m} x_{jk}}}$$

Então o algoritmo EM:

- Inicia ${P_1}^0$ e ${P_2}^0$

E-Step: determina (α') M-Step: encontra a solução $P_1^{(t+1)}$ e $P_2^{(t+1)}$ através de $(*)$, $(**)$, $(* * *)$.

3. De (A) temos que:

$$\log f_x(x|\theta) = \log f_y(y|\theta) - \log f_{z|x}(z|x, \theta)$$

Assim:

$$E\{\log f_x(x|\theta|x,\theta^{(t)})\} = E\{\log f_y(y|\theta)|x,\theta^{(t)}\} - E\{\log f_{z|x}(z|x,\theta)|x,\theta^{(t)}\} \; ... \; (4)$$

Onde as esperanças são calculadas com relação a $z|x,\theta^{(t)}$.
Logo, de (4) temos que:

$$\log f_x(x|\theta) = Q(\theta|\theta^{(t)}) - H(\theta|^{(t)}) \; ... \; (5)$$

Onde:

$$H(\theta|\theta^{(t)}) = E\{\log f_y(y|\theta)|x,\theta^{(t)}\}$$

A importância de 1 está relacionada a que $H(\theta|\theta^{(t)})$ é maximizada com relação a θ quando $\theta = \theta^{(t)}$

$$\begin{aligned} &H(\theta^{(t)}|\theta^{(t)}) - H(\theta|\theta^{(t)}) \\ &= E\{\log f_{z|x}(z|x,\theta^{(t)})| - \log f_{z|x}(z|x,\theta)\} \\ &= E\{-\log \left[\frac{f_{z|x}(z|x,\theta)}{f_{z|x,\theta^{(t)}}}\right] |x,\theta^{(t)}\} \\ &= \int -\log \left[\frac{f_{z|x}(z|x,\theta)}{f_{z|x,\theta^{(t)}}}\right] \times f_{z|x}(z|x,\theta^{(t)})dz \\ &\geq -\log \int f_{z|x}(z|x,\theta^{(t)})dz = 0 \end{aligned}$$

O último resultado é obtido da desigualdade de Jensen, já que $-\log u$ é estritamente convexa.

Se $\theta \neq \theta^{(t)} \rightarrow H(\theta|\theta^{(t)})$ é menor que $H(\theta^{(t)}|\theta^{(t)})$. Em particular se $O^{(t+1)}$ maximizar $Q(\theta|\theta^{(t)})$ com relação a θ,

$$\begin{aligned} &\log f_x(x|\theta^{(t+1)}) - \log f_x(x|\theta^{(t)}) \geq 0 \\ &\Longrightarrow L(\theta^{(t+1)}) \geq L(\theta^{(t)}) \end{aligned}$$

Já que Q cresce quando H decresce.

$$\Longrightarrow Q(\theta^{(t+1)}|\theta^{(t)}) > Q(\theta^{(t)}|\theta^{(t)})$$