Algoritmo EM: Parte 2 (extensões)

Prof. Caio Azevedo

Extensões do algoritmo EM

- Sejam $p(\mathbf{y}|\boldsymbol{\theta}) = L(\boldsymbol{\theta}, \mathbf{y})$ e $l(\boldsymbol{\theta}, \mathbf{y})$ a verossimilhança e a log-verossimilhança originais (com os dados incompletos).
- Sejam $p(\mathbf{y}, \mathbf{y}^*|\boldsymbol{\theta}) = L(\boldsymbol{\theta}, \mathbf{y}, \mathbf{y}^*)$ e $l(\boldsymbol{\theta}, \mathbf{y}, \mathbf{y}^*)$ a verossimilhança e a log-verossimilhança aumentadas.

Distribuição a posteriori

- $p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\Theta} p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$.
- $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{y}^*) = \frac{p(\mathbf{y},\mathbf{y}^*|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\Theta} p(\mathbf{y},\mathbf{y}^*|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto p(\mathbf{y}, \mathbf{y}^*|\boldsymbol{\theta})p(\boldsymbol{\theta})$.
- Assim, para encontrar a moda a posteriori (MAP) basta maximizar $\ln p(\mathbf{y}|\boldsymbol{\theta}) + \ln p(\boldsymbol{\theta})$ ou $\mathcal{E}(\ln p(\mathbf{y}, \mathbf{y}^*|\boldsymbol{\theta})|\mathbf{y}, \boldsymbol{\theta}^{(t)}) + \ln p(\boldsymbol{\theta})$.

Estrutura Bayesiana do algoritmo EM

- Seja $p(\boldsymbol{\theta}|\mathbf{y},\mathbf{y}^*)$ a log-posteriori aumentada e $\boldsymbol{\theta}^{(t)}$ estimativas provisórias para $\boldsymbol{\theta}$. O algoritmo EM pode ser resumido nos seguintes passos

Passo E: Calcule a esperança condicional (na log-posteriori) dos dados faltantes condicionado as variáveis observadas e à estimativas provisórias de $\boldsymbol{\theta}^{(t)}$, ou seja

$$Q_B(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathcal{E}(\ln p(\mathbf{y},\mathbf{y}^*|\boldsymbol{\theta})|\mathbf{y},\boldsymbol{\theta}^{(t)}) + \ln p(\boldsymbol{\theta})$$

Passo M: Maximizar a esperança acima em relação à $\boldsymbol{\theta}$ ou seja, obter

$$\boldsymbol{\theta}^{(t+1)} = \text{argmax}_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$$

até que que algum critério de convergência seja alcançado.

Extensões do algoritmo EM

- O Algoritmo EM é eficiente (produz boas estimativas) e flexível (pode ser aplicado em várias situações).
- O passo E e/ou o passo M podem ser difíceis de serem implementados.
- Passo E: integrais sem solução analítica.
- Passo M: equações sem solução analítica.
- Pode-se utilizar algum dos métodos de integração numérica para contornar o problema relacionado ao Passo E.
- Pode-se utilizar algum dos métodos de maximização numérica para contornar o problema relacionado ao Passo M.

Algumas extensões

- MCEM : Algoritmo EM via Monte Carlo.
- SEM: Algoritmo EM estocástico.
- CADEM: Algoritmo EM condicional via dados aumentados.
- ECM: Algoritmo EM com maximização condicional.
- PX-EM: Algoritmo EM via expansão paramétrica.
- Em princípio, qualquer uma das extensões pode ser estendida para a obtanção da moda a posteriori.

Estrutura do algoritmo MCEM

- Seja $l(\boldsymbol{\theta}, \mathbf{y}, \mathbf{y}^*)$ a log-verossimilhança aumentada e $\boldsymbol{\theta}^{(t)}$ estimativas provisórias para $\boldsymbol{\theta}$. O algoritmo EM pode ser resumido nos seguintes passos

Passo E: Calcule a esperança condicional (na log-verossimilhança) dos dados faltantes condicionado as variáveis observadas e à estimativas provisórias de $\boldsymbol{\theta}^{(t)}$, via integração por Monte Carlo, ou seja, simule m=1,...,M conjuntos de valores para $\mathbf{Y}^*|\mathbf{y}, \boldsymbol{\theta}^{(b)}$ e calcule

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathcal{E}\left[l(\boldsymbol{\theta}, \mathbf{y}, \mathbf{y}^*)|\mathbf{y}, \boldsymbol{\theta}\right] \approx \frac{1}{M}\sum_{i=1}^{M} l(\boldsymbol{\theta}, \mathbf{y}, \mathbf{y}_i^*)$$

Passo M: Maximizar a esperança acima em relação à $\boldsymbol{\theta}$ ou seja, obter

$$\boldsymbol{\theta}^{(t+1)} = \text{argmax}_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$$

Estrutura do algoritmo SEM (pode ser estendido para uma partição de ordem p das variáveis não observáveis)

- Considere $\mathbf{Y}^* = (\mathbf{Y}_1^*, \mathbf{Y}_2^*)$ em que $\mathbf{Y}_i^*$ e $\mathbf{Y}_i^*$ são diferentes tipos de dados não observadas (dados aumentados e efeitos aleatórios, por exemplo).

Passo S: (substituto do passo E). Simule valores para as variáveis não observadas através de $\mathbf{Y}_1^*|\mathbf{y}, \mathbf{y}_2^*, \boldsymbol{\theta}^{(t)}$ e $\mathbf{Y}_2^*|\mathbf{y}, \mathbf{y}_1^*, \boldsymbol{\theta}^{(t)}$ e impute esses valores na verossimilhança aumentada $l(\boldsymbol{\theta}, \mathbf{y}, \mathbf{y}^*)$.

Passo M: Maximizar a log-verossimilhança imputada acima em relação à $\boldsymbol{\theta}$ ou seja, obter

$$\boldsymbol{\theta}^{(t+1)} = \text{argmax}_{\boldsymbol{\theta}} l(\boldsymbol{\theta}, \mathbf{y}, \mathbf{y}^*)$$

até que que algum critério de convergência seja alcançado.

Estrutura do algoritmo CADEM (pode ser estendido para uma partição de ordem p das variáveis não observáveis)

- Considere $\mathbf{Y}^* = (\mathbf{Y}_1^*, \mathbf{Y}_2^*)$ em que $\mathbf{Y}_i^*$ e $\mathbf{Y}_i^*$ são diferentes tipos de dados não observadas (dados aumentados e efeitos aleatórios, por exemplo).

Passo E: . Calcule as esperançar condicionais das variáveis não observadas através de $\mathcal{E}(\mathbf{Y}_1^*|\mathbf{y}, \mathbf{y}_2^*, \boldsymbol{\theta}^{(t)})$ e $\mathcal{E}(\mathbf{Y}_2^*|\mathbf{y}, \mathbf{y}_1^*, \boldsymbol{\theta}^{(t)})$ e impute esses valores na verossimilhança aumentada $l(\boldsymbol{\theta}, \mathbf{y}, \mathbf{y}^*)$.

Passo M: Maximizar a log-verossimilhança imputada acima em relação à $\boldsymbol{\theta}$ ou seja, obter

$$\boldsymbol{\theta}^{(t+1)} = \text{argmax}_{\boldsymbol{\theta}} l(\boldsymbol{\theta}, \mathbf{y}, \mathbf{y}^*)$$

até que que algum critério de convergência seja alcançado.

Estrutura do algoritmo ECM (pode ser estendido para uma partição de ordem p dos parâmetros)

- Considere $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ em que $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$.

Passo E: Calcule a esperança condicional (na log-verossimilhança) dos dados faltantes condicionado as variáveis observadas e à estimativas provisórias de $\boldsymbol{\theta}^{(t)}$, ou seja

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathcal{E}\left[l(\boldsymbol{\theta}, \mathbf{y}, \mathbf{y}^*)|\mathbf{y}, \boldsymbol{\theta}\right]$$

Passo CM: (substituto do passo M) Maximizar $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ da seguinte forma $\boldsymbol{\theta}_1^{(t+1)} = \text{argmax}_{\boldsymbol{\theta}_1} Q(\boldsymbol{\theta}|\boldsymbol{\theta}_2^{(t)}), \boldsymbol{\theta}_2^{(t+1)} = \text{argmax}_{\boldsymbol{\theta}_2} Q(\boldsymbol{\theta}|\boldsymbol{\theta}_1^{(t+1)})$

até que que algum critério de convergência seja alcançado.