

(1001919) Métodos computacionalmente intensivos

Lista de fixação 5

Exercício 1. Um pesquisador prepara $n = 30$ placas Petri com um substrato. Considere que X_i é a quantidade de substrato aplicada na i -ésima placa. Após 24h, o pesquisador conta o número de bactérias que se desenvolveram em cada placa. Considere que Y_i é o número de bactérias contadas na i -ésima placa. O pesquisador acredita que o número de bactérias contado deve ser proporcional à quantidade de substrato disponível. Assim, define que $Y_i|X_i \sim \text{Poisson}(\theta \cdot X_i)$.

- Simule dados, (X_i, Y_i) , tomando $\theta = 0.5$ e $X_i \sim \text{Exp}(1)$.
- Determine a log-verossimilhança, a função score e a matriz de informação de Fisher no modelo proposto.
- Utilize o método escore de Fisher para determinar o estimador de máxima verossimilhança para θ com base nos dados simulados.

Exercício 2. Considere que X_1, \dots, X_n são tais que $X_i \in \mathbb{R}$. A análise de agrupamento usando k -médias muitas vezes é realizada utilizando o algoritmo de Lloyd, em que é escolhido uma inicialização, $\theta^{(0)}$, e seguem-se as seguintes iterações para cada $1 \leq j \leq B$:

- Para cada $1 \leq i \leq n$, Calcule $Z_{i,k}^{(j)} = \mathbb{I}(|\theta_k^{(j-1)} - x_i| < |\theta_{1-k}^{(j-1)} - x_i|)$
- Calcule $\theta_k^{(j)} = \frac{\sum_{i=1}^n Z_{i,k}^{(j)} X_i}{\sum_{i=1}^n Z_{i,k}^{(j)}}$.

Com base nesta exposição, resolva as seguintes questões:

- Gere dados X_1, \dots, X_n tais que $Z_i \sim \text{Bernoulli}(0.5)$ e $X_i|Z_i \sim N(\theta_i, 1)$, onde $(\theta_0, \theta_1) = (-0.5, 0.75)$ e $n = 100$.
- Implemente o algoritmo de Lloyd e aplique-o nos dados anteriormente gerados para $B = 10^4$. Indique a estimativa obtida para θ , isto é, $(\theta_0^{(B)}, \theta_1^{(B)})$.
- Considere o modelo de mistura de normais em que $Z_i \sim \text{Bernoulli}(p)$ e $X_i|Z_i \sim N(\theta_Z, \sigma)$, com p e σ conhecidos. Prove que $\lim_{\sigma \rightarrow 0} \mathbb{P}(Z_i = 1|X_i = x_i, \theta) = \mathbb{I}(|\theta_1 - x_i| < |\theta_0 - x_i|)$. Argumente que o algoritmo EM obtido neste modelo é equivalente ao algoritmo de Lloyd quando $\sigma \rightarrow 0$.
- Aplique o algoritmo EM nos dados gerados baseado no modelo de mistura de normais com $p = 0.5$ e $\sigma = 1$. Compare a estimativa para θ com aquela obtida pelo algoritmo de Lloyd.

Exercício 3. Considere que $Z_i \sim \text{Bernoulli}(0.5)$ e $X_i|Z_i \sim \text{Exp}(\theta_{Z_i})$.

- Gere X_1, \dots, X_{100} com base no modelo proposto usando $\theta_0 = 1$ e $\theta_1 = 3$.
- Implemente o algoritmo EM para determinar o estimador de máxima verossimilhança para θ baseado em X . Indique a estimativa para θ .

Referências