

Exemplo de Mistura Gaussiana para Classificação por Altura

Enunciado da Questão

Considere uma população de 5 pessoas com as seguintes alturas (em cm): 165, 182, 168, 178 e 171. Você deve usar um modelo de mistura gaussiana (GMM) para separar essa população em dois grupos distintos com base na altura. Interprete os grupos resultantes como "adolescentes" e "adultos".

1 Dados do Problema

Considere uma população de 5 pessoas com as seguintes alturas (em cm):

- Pessoa 1: 165
- Pessoa 2: 182
- Pessoa 3: 168
- Pessoa 4: 178
- Pessoa 5: 171

Queremos usar um modelo de mistura gaussiana (GMM) para separar essa população em 2 grupos distintos baseados na altura, interpretando-os como "adolescentes" e "adultos".

2 Passo 1: Inicialização dos Parâmetros

Inicializamos os seguintes parâmetros para os grupos de adolescentes e adultos:

2.1 Parâmetros para Adolescentes (Grupo 1)

- Peso de mistura: $\pi_1 = 0.5$
- Média: $\mu_1 = 167$ cm
- Variância: $\sigma_1^2 = 25$ cm²

2.2 Parâmetros para Adultos (Grupo 2)

- Peso de mistura: $\pi_2 = 0.5$
- Média: $\mu_2 = 180$ cm
- Variância: $\sigma_2^2 = 25$ cm²

3 Passo 2: Algoritmo EM (Expectation-Maximization)

3.1 Iteração 1

3.1.1 Passo E (Expectation): Cálculo das Responsabilidades

Para cada altura x_i , calculamos a probabilidade de pertencer a cada grupo:

$$p(x_i|k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right) \quad (1)$$

$$\gamma(i, k) = \frac{\pi_k \cdot p(x_i|k)}{\pi_1 \cdot p(x_i|1) + \pi_2 \cdot p(x_i|2)} \quad (2)$$

Cálculos para a Pessoa 1 ($x_1 = 165$):

$$p(x_1|1) = \frac{1}{\sqrt{50\pi}} \exp\left(-\frac{(165 - 167)^2}{50}\right) = 0.0736$$

$$p(x_1|2) = \frac{1}{\sqrt{50\pi}} \exp\left(-\frac{(165 - 180)^2}{50}\right) = 0.0009$$

$$\gamma(1, 1) = \frac{0.5 \cdot 0.0736}{0.5 \cdot 0.0736 + 0.5 \cdot 0.0009} = 0.9877$$

$$\gamma(1, 2) = \frac{0.5 \cdot 0.0009}{0.5 \cdot 0.0736 + 0.5 \cdot 0.0009} = 0.0123$$

Cálculos para a Pessoa 2 ($x_2 = 182$):

$$p(x_2|1) = \frac{1}{\sqrt{50\pi}} \exp\left(-\frac{(182 - 167)^2}{50}\right) = 0.0009$$

$$p(x_2|2) = \frac{1}{\sqrt{50\pi}} \exp\left(-\frac{(182 - 180)^2}{50}\right) = 0.0736$$

$$\gamma(2, 1) = \frac{0.5 \cdot 0.0009}{0.5 \cdot 0.0009 + 0.5 \cdot 0.0736} = 0.0123$$

$$\gamma(2, 2) = \frac{0.5 \cdot 0.0736}{0.5 \cdot 0.0009 + 0.5 \cdot 0.0736} = 0.9877$$

Cálculos para a Pessoa 3 ($x_3 = 168$):

$$p(x_3|1) = \frac{1}{\sqrt{50\pi}} \exp\left(-\frac{(168 - 167)^2}{50}\right) = 0.0782$$

$$p(x_3|2) = \frac{1}{\sqrt{50\pi}} \exp\left(-\frac{(168 - 180)^2}{50}\right) = 0.0045$$

$$\gamma(3, 1) = \frac{0.5 \cdot 0.0782}{0.5 \cdot 0.0782 + 0.5 \cdot 0.0045} = 0.9450$$

$$\gamma(3, 2) = \frac{0.5 \cdot 0.0045}{0.5 \cdot 0.0782 + 0.5 \cdot 0.0045} = 0.0550$$

Cálculos para a Pessoa 4 ($x_4 = 178$):

$$p(x_4|1) = \frac{1}{\sqrt{50\pi}} \exp\left(-\frac{(178 - 167)^2}{50}\right) = 0.0071$$

$$p(x_4|2) = \frac{1}{\sqrt{50\pi}} \exp\left(-\frac{(178 - 180)^2}{50}\right) = 0.0736$$

$$\gamma(4, 1) = \frac{0.5 \cdot 0.0071}{0.5 \cdot 0.0071 + 0.5 \cdot 0.0736} = 0.0876$$

$$\gamma(4, 2) = \frac{0.5 \cdot 0.0736}{0.5 \cdot 0.0071 + 0.5 \cdot 0.0736} = 0.9124$$

Cálculos para a Pessoa 5 ($x_5 = 171$):

$$p(x_5|1) = \frac{1}{\sqrt{50\pi}} \exp\left(-\frac{(171 - 167)^2}{50}\right) = 0.0579$$

$$p(x_5|2) = \frac{1}{\sqrt{50\pi}} \exp\left(-\frac{(171 - 180)^2}{50}\right) = 0.0158$$

$$\gamma(5, 1) = \frac{0.5 \cdot 0.0579}{0.5 \cdot 0.0579 + 0.5 \cdot 0.0158} = 0.7857$$

$$\gamma(5, 2) = \frac{0.5 \cdot 0.0158}{0.5 \cdot 0.0579 + 0.5 \cdot 0.0158} = 0.2143$$

4 Resultados após a Primeira Iteração

Cada valor de $\gamma(i, k)$ representa a probabilidade de a altura x_i pertencer ao grupo k .

5 Passo 3: Classificação Final

Para classificar cada pessoa, comparamos as responsabilidades:

- Pessoa 1 (165 cm): Grupo 1 ($\gamma(1, 1) = 0.9877 > \gamma(1, 2) = 0.0123$)

- Pessoa 2 (182 cm): Grupo 2 ($\gamma(2, 2) = 0.9877 > \gamma(2, 1) = 0.0123$)
- Pessoa 3 (168 cm): Grupo 1 ($\gamma(3, 1) = 0.9450 > \gamma(3, 2) = 0.0550$)
- Pessoa 4 (178 cm): Grupo 2 ($\gamma(4, 2) = 0.9124 > \gamma(4, 1) = 0.0876$)
- Pessoa 5 (171 cm): Grupo 1 ($\gamma(5, 1) = 0.7857 > \gamma(5, 2) = 0.2143$)

6 Conclusão

A separação em dois grupos ficou:

- Grupo 1 (adolescentes): Pessoas 1, 3 e 5 com alturas 165 cm, 168 cm e 171 cm
- Grupo 2 (adultos): Pessoas 2 e 4 com alturas 182 cm e 178 cm

Em um caso real, continuaríamos as iterações até a convergência, mas já nesta primeira iteração conseguimos observar uma separação clara dos grupos.