

Aprendizado de Máquina e Reconhecimento de Padrões (UTFPR/CPGEI) - Lista de Exercícios 5

Tópicos: Clustering.

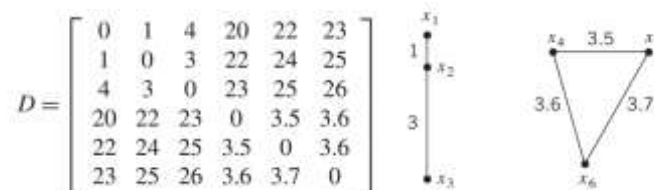
1. Gere um conjunto de dados bidimensionais \mathbf{X} com 400 exemplos. Esses exemplos formam quatro grupos igualmente distribuídos seguindo distribuições Gaussianas dadas pelos parâmetros:

$$m_1 = [0, 0]^T, m_2 = [10, 0], m_3 = [0, 9], \text{ and } m_4 = [9, 8]^T$$

$$S_1 = I, \quad S_2 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1.5 \end{bmatrix}, \quad S_3 = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1.1 \end{bmatrix}, \quad S_4 = \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.5 \end{bmatrix}$$

- a) Aplique o modelo k-means com $m = 4$ (número de centros). Utilize a função “rand” do MATLAB para inicializar os parâmetros do k-means. Compare os centros obtidos pelo k-means com as médias das Gaussianas acima apresentadas. Plote os centros do k-means e os parâmetros \mathbf{m} das Gaussianas e compare as diferenças.
 - b) Repita o processo (a) para $m = 3$.
 - c) Repita o processo (a) para $m = 5$.
 - d) Repita o processo (a), agora inicializando o modelo k-means com $[-2.0, -2.0]^T$, $[-2.1, -2.1]^T$, $[-2.0, -2.2]^T$, $[-2.1, -2.2]^T$.
 - e) Repita o processo (a), inicializando os três primeiros centros do k-means aleatoriamente (rand) e o quarto centro com $[20, 20]^T$.
 - f) Repita o processo (a), (b), (c) e (d) utilizando o modelo Fuzzy C-Means e compare com os resultados obtidos para o k-Means.
 - g) Comente os resultados.
2. Gere um conjunto de dados bidimensionais \mathbf{X} com 400 pontos situados em torno de dois círculos concêntricos. Os dois primeiros 200 devem estar em torno de um círculo de raio 3 centrado em $(0, 0)$. Os demais devem estar em torno de um círculo de raio 6 centrado em $(1, 1)$. Aplique um algoritmo de clustering espectral com medida de similaridade Gaussiana considerando $\epsilon=1.5$ e $\sigma=2$ e plote os resultados. Altere ϵ e σ e rode novamente. Qual o efeito de ϵ e σ nos resultados?

3. Considere o conjunto de dados $\{x_1, x_2, x_3, x_4, x_5, x_6\}$, sendo:



A matriz que representa as distâncias d_{ij} entre pares de pontos entre os vetores \mathbf{x}_i e \mathbf{x}_j , conforme ilustrado na figura ao lado da matriz. Aplique, o algoritmo single-link e avalie o resultado. Repita o processo usando o complete-link. Apresente ambos os dendrogramas.