Inteligência Computacional

Professor: Prof. Dr. Rogério Martins Gomes. Aluno: Rodrigo Rodrigues de Novaes Júnior.
Terceira lista de exercícios, 22 de janeiro de 2018.

Questão 04

O método Elbow (cotovelo) é bastante utilizado para essa finalidade. Consiste em testar os possíveis valores de K, partindo de 1, até que a variação da função custo seja tão pequena que possa ser desprezada. Isso faz com que exista um ponto, conhecido como "cotovelo", cuja variação para o ponto anterior é elevada, mas pequena em relação ao sucessor. O método indica que esse valor seja adequado para atribuir a K.

Outra estratégia plausível é factível busca binária com base na minimização da função custo. Se escolhermos um intervalo $K \in [1, L)$ tal que a função custo $J(\ldots)_K$ seja estritamente decrescente, pode-se executar o seguinte algoritmo:

$$find(l_i, l_f) = \begin{cases} find\left(l_i, \dfrac{l_i + l_f}{2}\right), & \text{se } J(\ldots)_{(l_i+l_f)/2)} \geq J(\ldots)_{l_i} \\ find\left(\dfrac{l_i + l_f}{2}, l_f\right), & \text{se } J(\ldots)_{(l_i+l_f)/2)} < J(\ldots)_{l_i} \\ \dfrac{l_i + l_f}{2}, & \text{se } l_i = l_f + 1, \end{cases}$$

onde $J(\ldots)_K$ é a função custo avaliada para um valor de K, l_i o valor inferior e l_f o valor superior de um subintervalo pertencente a $[1, L)$, tal que para todo $[l_i, l_f) \subseteq [1, L)$, o valor de K deve estar contido em $[l_i, l_f)$. Nessas condições, $K = find(1, L)$ fará com que $J(\ldots)_K$ seja mínimo.

Questão 06

Seja x_1 o vetor que representa a entrada em ml/dia de café e x_2 a entrada em ml/dia de leite, sabendo que precisamos gerar $K = 3$ perfis de pessoas, bem como tendo os seguintes centroides iniciais:

$$C_1(10, 30); C_2(45, 46); C_3(55, 57),$$

precisamos utilizar a distância euclidiana, dada por

$$d((x_a, y_a), (x_b, y_b)) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \tag{1}$$

para definir as novas posições de C_1, C_2 e C_3 a partir das médias de todo $p_j \in P$, $P = [x_1 x_2]$ sendo a matriz de entrada, onde i representa um dos agrupamentos. Nesse contexto, a seguinte

tabela apresenta o agrupamento de uma entrada e a distância euclidiana para seu centroide correspondente:

i	p_j	$d(C_i, p_j)$
1	(32, 27)	22.203603
2	(55, 43)	10.440307
3	(80, 63)	25.709920
3	(85, 50)	30.805844
2	(58, 38)	15.264338
3	(82, 55)	27.073973
1	(25, 31)	15.033296
3	(66, 42)	18.601075
3	(60, 49)	9.433981
1	(35, 12)	30.805844

Isso responde ao item **a)** da questão. Para o item **b)**, define-se um novo centroide por

$$C_i' \left(\frac{1}{|i|} \sum_{j \in C_i} x_{1j}, x_{2j}) \right).$$

onde $|i|$ é o número de elementos no agrupamento i. Portanto, os novos centroides são:

$$C_1(30.67, 23.33); C_2(56.50, 40.50); C_3(74.60, 51.80).$$

Ao recalcular o agrupamento de cada p_j, temos:

i	p_j	$d(C_i, p_j)$
1	(32, 27)	3.903562
2	(55, 43)	2.915476
3	(80, 63)	12.433825
3	(85, 50)	10.554620
2	(58, 38)	2.915476
3	(82, 55)	2.915476
1	(25, 31)	9.538228
2*	(66, 42)	9.617692
2*	(60, 49)	9.192388
1	(35, 12)	12.129213

Todas as linhas marcadas com * apresentaram mudanças de agrupamento. A tabela final é mostrada acima.

Questão 07

Queremos calcular

$$\Sigma = \frac{1}{m} \sum_{i=1}^{m} \left(x^{(i)}\right) \left(x^{(i)}\right)^T,$$

onde m é a dimensão da entrada e A' é a transposta de A.