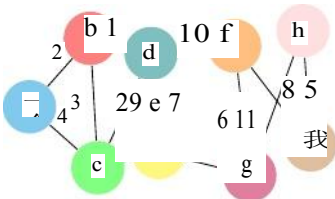


1.对以下输入执行 k-中值  $\mu_1 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$   $\mathbf{x}_2 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$   $\mathbf{x}_3 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$   $\mathbf{x}_4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$   $\mathbf{x}_5 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ .  
假设  $k=2$ ， 集群的初始集群位置为

$$\mu_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \mu_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}.$$

记住， 两点之间的距离是两个绝对值的和。例如， 之间的距离  $x_1$  和  $x_2$  是 $|3-4| + |4-2| = 3$ 。如果出现平局， 则将数据点分配给 ID 较小的集群。  
2.如果我们使用 k-Means， 第一个任务和更新步骤会是什么样子？

3.在下面的相似图上执行完全链接。请注意， 所有不存在的边的相似性都是 0!还要注意  $\text{sim}(f, g) = 6$ ,  $\text{sim}(g, h) = 11$ ,  $\text{sim}(f, i) = 8$ 。回想一下， 在出现 tie 的情况下， 总是选择 ID 最小的集群， 然后将其与 ID 最小的集群合并。这里最小的 ID 是按字母顺序给出的。例如， 假设你有  $S_1 = \{a, c\}$ ，  $S_2 = \{b, f\}$ ，  $S_3 = \{d, h\}$ ，  $\text{sim}(S_1, \text{年代}_2) = 4$ ,  $\text{sim}(S_1, \text{年代}_3) = 5$  和  $\text{sim}(S_2, \text{年代}_3) = 5$ 。这里有个平局， 是否合并  $S_1$  和  $S_3$  或年代  $_2$  和  $S_3$ ? S 的 ID<sub>1</sub> 是 a, S 的 ID<sub>2</sub> b 和 S 的 ID 是多少  $_3$  是 d， 所以我们选  $S_1$  然后我们将它与所有 tie 中 ID 最小的集合合并。在本例中， 只有  $S_3$ 。我们合并  $S_1$  和  $S_3$ 。  
这是图表， 祝你好运!



第 6 课:答案

(版本 1.0)

1.对以下输入执行 k-中值  $\mu_1 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$   $\mathbf{x}_2 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$   $\mathbf{x}_3 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$   $\mathbf{x}_4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$   $\mathbf{x}_5 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ .  
假设  $k=2$ ，集群的初始集群位置为

$$\mu_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \mu_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}.$$

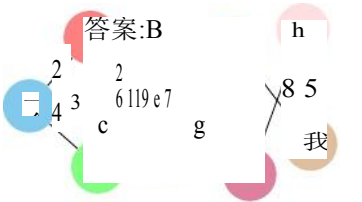
记住，两点之间的距离是两个绝对值的和。例如， $\mathbf{x}_1$  和  $\mathbf{x}_2$  之间的距离是  $|3-4| + |4-2| = 3$ 。如果出现平局，则将数据点分配给 ID 较小的集群。解决方案:

- (a)赋值 1:  $S_1 = \{\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$ ,  $S_2 = \{\mathbf{x}_1\}$ 。
- (b)更新 1:  $\mu_1 = \begin{pmatrix} 7/4 \\ 2 \end{pmatrix}$  and  $\mu_2 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ 。
- (c)赋值 2:  $S_1 = \{\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$ ,  $S_2 = \{\mathbf{x}_1\}$ 。
- (d)不变，算法终止

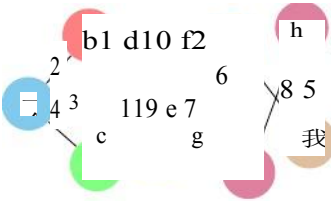
2.如果我们使用 k-Means，第一个任务和更新步骤会是什么样子?解决方案:

- (a)赋值 1:  $S_1 = \{\mathbf{x}_3, \mathbf{x}_4\}$ ,  $S_2 = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5\}$ 。
- (b)更新 1:  $\mu_1 = \begin{pmatrix} 1 \\ 1.5 \end{pmatrix}$  and  $\mu_2 = \begin{pmatrix} 8/3 \\ 3 \end{pmatrix}$ 。

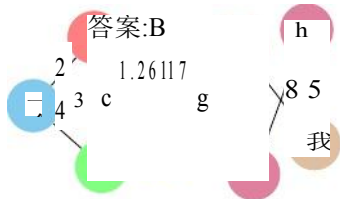
3.在下面的相似图上执行完全链接。请注意，所有不存在的边的相似性都是 0!还要注意  $\text{sim}(\mathbf{f}, \mathbf{g}) = 6$ ,  $\text{sim}(\mathbf{g}, \mathbf{h}) = 11$ ,  $\text{sim}(\mathbf{f}, \mathbf{i}) = 8$ 。回想一下，在出现 tie 的情况下，总是选择 ID 最小的集群，然后将其与 ID 最小的集群合并。这里最小的 ID 是按字母顺序给出的。例如，假设你有  $S_1 = \{\mathbf{c}\}$ ,  $S_2 = \{\mathbf{b}, \mathbf{f}\}$ ,  $S_3 = \{\mathbf{d}, \mathbf{h}\}$  with  $\text{sim}(S_1, S_2) = 4$ ,  $\text{sim}(S_1, S_3) = 5$  and  $\text{sim}(S_2, S_3) = 5$ . So we have a tie 这里:我们是否合并  $S_1$  和  $S_3$  或年代  $S_2$  和  $S_3$ ?  $S_1$  的 ID 是 a,  $S_2$  的 ID 是 b 和  $S_3$  的 ID 是多少 c 是 d, 所以我们选  $S_1$  然后将它与所有 tie 中 ID 最小的集合合并。在本例中，只有  $S_3$ 。我们合并  $S_1$  和  $S_3$ 。  
这是图表，祝你好运!



•第一轮:



•第二轮:

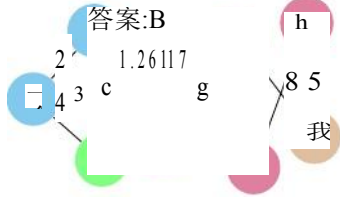


•第三轮:

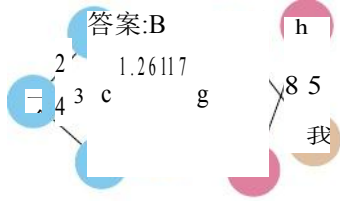


•轮 4:

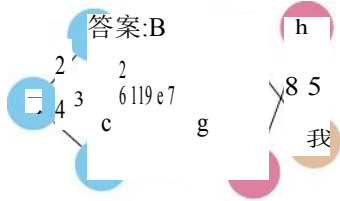
注意，不存在的边的值为 0。Sim ({d, f}, {i})=0 表示完全连锁。同样，sim({c, e}, {g}), sim({d, f}, {g, h}), sim({h, g}, {i})以及其他所有边，直到我们到达 a 和 b 之间的边。



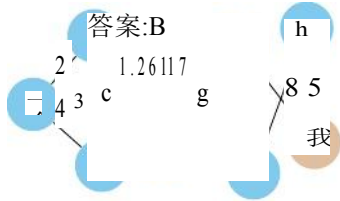
•轮 5:



•轮 6:



•轮 7:



•轮 8:

