19.01.2022

Assignment 8: "Clustering and Modularity"

Problem 8-2 — Hierarchical clustering

1. Single-linkage and complete-linkage swap their definition when used with a similarity rather than a distance function due to their (usual) relation between each other. According to the usual definitions the similarity function has an inverse relation to the distance function, in other words: nodes are more similar when their distance is smaller and vice versa.

A maximum of the distance function therefore corresponds to a minimum of the similarity function, which is why "min" and "max" are swapped.

2.
$$
x_{ij}^{0} = \begin{array}{c} \begin{pmatrix} 0 & 1 & \frac{1}{2} & 1 & 1 & \frac{1}{2} \\ 1 & 0 & \frac{1}{3} & 1 & 1 & 0 \\ \frac{1}{2} & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 1 \\ 1 & 1 & \frac{1}{3} & 0 & 1 & 0 \\ 1 & 1 & \frac{1}{3} & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{array}{l} \to A \\ \to B \\ \vdots \\ \vdots \\ \vdots \\ \to F \end{array} \\ \begin{array}{cccccc} \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ A & B & C & D & E & F \end{array} \end{array}
$$

Let the nodes of this network be $\{A, B, C, D, E, F\}$

Agglomerative hierarchical clustering using single-linkage

$\boxed{\text{Step 0}}$ clusters: $C_0^{(0)} = \{A\}, C_1^{(0)} = \{B\}, C_2^{(0)} = \{C\}, C_3^{(0)} = \{D\}, C_4^{(0)} = \{E\}, C_5^{(0)} = \{F\}$

a) maximum similarity: $\max\limits_{x \in X, y \in Y} s(X,Y) = 1$ (X, Y disjoint clusters)

e.g. for clusters $(C_0^{(0)}, C_1^{(0)})$, $(C_0^{(0)}, C_3^{(0)})$, $(C_0^{(0)}, C_4^{(0)})$, ...

b) merge clusers $C_0^{(1)} \leftarrow C_0^{(0)} \cup C_1^{(0)} = \{A, B\}$

c) update similarity matrix:

$$
x_{ij}^{(1)} = \begin{array}{c} \begin{pmatrix} 0 & \frac{1}{2} & 1 & 1 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{3} & \frac{1}{3} & 1 \\ 1 & \frac{1}{3} & 0 & 1 & 0 \\ 1 & \frac{1}{3} & 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 & 0 & 0 \end{pmatrix} \\ \begin{array}{ccccc} \downarrow & \downarrow & & & \downarrow \\ C_0^{(1)} & C_1^{(1)} & \dots & & C_4^{(1)} \end{array} \end{array}
$$

e.g. $x_{01}^{(1)} = \max(x_{02}^{0}, x_{12}^{0})$
$= \max(\frac{1}{2}, \frac{1}{3})$
$= \frac{1}{2}$

...

$\boxed{\text{step 1}}$ clusters : $C_0^{(1)} = \{A,B\}$, $C_1^{(1)} = \{C\}$, $C_2^{(1)} = \{D\}$, $C_3^{(1)} = \{E\}$, $C_4^{(1)} = \{F\}$

a) max $S(X,Y) = 1$, e.g. for clusters $(C_1^{(1)}, C_4^{(1)})$

b) merge: $C_1^{(2)} \leftarrow C_1^{(1)} \cup C_4^{(1)} = \{C,F\}$

c) update:

$$x_{ij}^{(2)} = \begin{array}{c} \begin{pmatrix} 0 & 1/2 & 1 & 1 \\ 1/2 & 0 & 1/3 & 1/3 \\ 1 & 1/3 & 0 & 1 \\ 1 & 1/3 & 1 & 0 \end{pmatrix} \\ C_0^{(2)} \ldots\ldots C_3^{(2)} \end{array}$$

$\boxed{\text{step 2}}$ clusters : $C_0^{(2)} = \{A,B\}$, $C_1^{(2)} = \{C,F\}$, $C_2^{(2)} = \{D\}$, $C_3^{(2)} = \{E\}$

a) max $S(X,Y) = 1$, e.g. for clusters $(C_2^{(2)}, C_3^{(2)})$

b) merge : $C_2^{(3)} \leftarrow C_2^{(2)} \cup C_3^{(2)} = \{D, E\}$

c) update :

$$x_{ij}^{(3)} = \begin{array}{c} \begin{pmatrix} 0 & 1/2 & 1 \\ 1/2 & 0 & 1/3 \\ 1 & 1/3 & 0 \end{pmatrix} \\ C_0^{(3)} \; C_1^{(3)} \; C_2^{(3)} \end{array}$$

$\boxed{\text{step 3}}$ clusters : $C_0^{(3)} = \{A,B\}$, $C_1^{(3)} = \{C,F\}$, $C_2^{(3)} = \{D,E\}$

a) max $S(X,Y) = 1$ for clusters $(C_0^{(3)}, C_2^{(3)})$

b) merge : $C_0^{(4)} \leftarrow C_0^{(3)} \cup C_2^{(3)} = \{A,B,D,E\}$

c) update:

$$x_{ij}^{(4)} = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}$$

$\boxed{\text{step 4}}$ clusters : $C_0^{(4)} = \{A,B,D,E\}$, $C_1^{(4)} = \{C,F\}$

a) max $S(X,Y) = 1/2$ for clusters $(C_0^{(4)}, C_1^{(4)})$

b) merge : $C_0^{(5)} \leftarrow C_0^{(4)} \cup C_1^{(4)} = \{A,B,C,D,E,F\}$

- end -

dendrogram