Exercises for Chapter 9

9.2 Consider the relational data set

$$D = \begin{pmatrix} 0 & 1 & 4 \\ 1 & 0 & 2 \\ 4 & 2 & 0 \end{pmatrix}.$$

a) Give the results of the c-medoids algorithm, $c = 2$, for all possible initializations of V.

$V = \{x_1, x_2\} : d_{13} > d_{23} \Rightarrow U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \Rightarrow V = \{x_1, x_2\} \vee \{x_1, x_3\}$,

$V = \{x_1, x_3\} : d_{12} < d_{23} \Rightarrow U = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \Rightarrow V = \{x_1, x_3\} \vee \{x_2, x_3\}$,

$V = \{x_2, x_3\} : d_{12} < d_{13} \Rightarrow U = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \Rightarrow V = \{x_1, x_3\} \vee \{x_2, x_3\}$

So, there are three possible stable results:

$V = \{x_1, x_2\}$, $U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ for initalization $V = \{x_1, x_2\}$,

$V = \{x_1, x_3\}$, $U = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ for any initalization,

$V = \{x_2, x_3\}$, $U = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ for any initalization

b) Why is relational c-means not suitable for this data set?

Because D is not Euclidean: $d_{12} + d_{23} < d_{13}$.

c) How could the data be transformed to be used by relational c-means?

Using the β spread transformation: $D' = \begin{pmatrix} 0 & 1+\beta & 4+\beta \\ 1+\beta & 0 & 2+\beta \\ 4+\beta & 2+\beta & 0 \end{pmatrix}$

d) Compute the transformed data set.

For the triangle inequality to hold, the minimum value of β is found by $d_{12} + d_{23} = d_{13} \quad \Rightarrow \quad 1 + \beta + 2 + \beta = 4 + \beta \quad \Rightarrow \quad \beta = 1$

$$\Rightarrow \quad D' = \begin{pmatrix} 0 & 2 & 5 \\ 2 & 0 & 3 \\ 5 & 3 & 0 \end{pmatrix}$$

e) Find a feature data set that corresponds to this transformed relational data set.

The data are collinear, so they can be one–dimensional, three points with distances D': $X' = \{0, 2, 5\}$ plus any real valued offset, or $X' = \{5, 3, 0\}$ plus any real valued offset.