25.11.17

Problem Set 5 MMD

Exercise 1

a)

	a	b	c	d	e	f	g	h
A	1	1	0	1	0	0	1	0
B	0	1	1	1	0	0	0	0
C	0	0	0	1	0	1	1	1

Jaccard distance matrix

	a	b	c	d	e	f	g	h
a	0	1/2	1	2/3	1	1	1/2	1
b		0	1/2	1/3	1	1	2/3	1
c			0	2/3	1	1	1	1
d				0	1	2/3	1/3	2/3
e					0	1	1	1
f						0	1/2	$\boxed{0}$
g							0	1/2
h								0

-> Min-Element

The Jaccard distance in the binary case is computed

$$d(A,B) = \frac{M_{11}}{M_{01} + M_{10} + M_{11}}$$

M_{11} - # attributes where A and B both have a value of 1

M_{10} - # attributes where A has 1 and B has 0

M_{01} - # attributes where A has 0 and B has 1.

* Clustering (using single linkage)

Step 1: We cluster according to the selected Min-Element f and h together (actually they could have been clustered beforehand as they are identical)

	a	b	c	d	e	fh	g
a	0	1/2	1	2/3	1	1	1/2
b		0	1/2	1/3	1	1	2/3
c			0	2/3	1	1	1
d				0	1	2/3	$\boxed{1/3}$
e					0	1	1
fh						0	1/2
g							0

-> select as Min Element

* Note: as there was no rule defined, I chose randomly when there were several Min Elements

Step 2: Cluster d and g

	a	b	c	dg	e	fh
a	0	1/2	1	1/2	1	1
b		0	1/2	1/3	1	1
c			0	2/3	1	1
dg				0	1	2/3
e					0	1
fh						0

→ Min Element

Step 3: Cluster b and dg

	a	c	bdg	e	fh
a	0	1	1/2	1	1
c		0	1/2	1	1
bdg			0	1	1/2
e				0	1
fh					0

→ select as Min Element

⇒ Final clusters: {(abdg), (c), (e), (fh)}

b) User-cluster matrix

users	abdg	c	e	fh
A	4.25	-	1	2
B	2.5	4	1	2
C	3.5	1	-	4.5

clusters

c) Cosine distances

$$\cos(A,B) = 1 - sim(A,B)$$
$$= 1 - \frac{4.25 \cdot 2.5 + 4 \cdot 0 + 1 \cdot 1 + 2 \cdot 2}{\sqrt{4.25^2 + 1^2 + 2^2} \cdot \sqrt{2.5^2 + 4^2 + 1^2 + 2^2}}$$
$$= 0.402$$

$$\cos(A,C) = 1 - sim(A,C)$$
$$= 1 - \frac{4.25 \cdot 3.5 + 0 \cdot 1 + 1 \cdot 0 + 2 \cdot 4.5}{\sqrt{4.25^2 + 1^2 + 2^2} \cdot \sqrt{3.5^2 + 1^2 + 4.5^2}}$$
$$= 0.152$$

$$\cos(B,C) = 1 - sim(B,C)$$
$$= 1 - \frac{2.5 \cdot 3.5 + 4 \cdot 1 + 1 \cdot 0 + 2 \cdot 4.5}{\sqrt{2.5^2 + 4^2 + 1^2 + 2^2} \cdot \sqrt{3.5^2 + 1^2 + 4.5^2}}$$
$$= 0.29$$