CI603 Data Mining

Tutorial 2 Part II

(solution)

1. Using the table below, plot the points on a graph. For each pair of points calculate the Euclidian, and Manhattan distance between them showing your results in a distance matrix.

	X1(x)	X2(y)
P1	2.3	4
P2	1.5	3.2
P3	4.2	2
P4	1.7	3.7

Euclidean distance

	P1	P2	P3	P4
P1	0.000000	1.131371	2.758623	0.670820
P2	1.131371	0.000000	2.954657	0.538516
P3	2.758623	2.954657	0.000000	3.023243
P4	0.670820	0.538516	3.023243	0.000000

Manhattan distance

	P1	P2	P3	P4
P1	0.0	1.6	3.9	0.9
P2	1.6	0.0	3.9	0.7
P3	3.9	3.9	0.0	4.2
P4	0.9	0.7	4.2	0.0

2. Using the table below calculate the Euclidean and Manhattan distance of the below data.

	Att1	Att2	Att3	Att4
Object 1	2.1	5.3	6.4	5.7
Object 2	1.4	2.7	2.3	3.2
Object 3	4.2	1.1	1.8	1.8
Object 4	2.1	2.2	2.9	7.2

Euclidean distance

	Object 1	Object 2	Object 3	Object 4
Object 1	0.000000	5.505452	7.643298e+00	4.910193
Object 2	5.505452	0.000000	3.551056e+00	4.135215
Object 3	7.643298	3.551056	1.192093e-07	5.999167
Object 4	4.910193	4.135215	5.999167e+00	0.000000

Manhattan distance

	Object 1	Object 2	Object 3	Object 4
Object 1	0.0	9.9	14.8	8.1
Object 2	9.9	0.0	6.3	5.8
Object 3	14.8	6.3	0.0	9.7
Object 4	8.1	5.8	9.7	0.0

3. For the following vectors, x and y calculate the indicated similarity or distance measure.

   a) x= (5,1,3,1), y = (0,2,0,2) cosine, correlation, Euclidean.

      cos(x,y) = **0.2357022603955159**
      corr(x,y) = **-0.9045**
      Euclidean (x,y) = **6.0**

b) x = (0,1,0,1), y = (1,0,1,0) cosine, correlation, Euclidean, Jaccard.

cos(x,y) = **0.0**

corr(x,y) = **-1.0**

Euclidean (x,y) = **2.0**

Jaccard (x,y) = **0.0**

c) x= (0,-1,0,1), y = (1,0,-1,0) cosine, correlation, Euclidean.

cos(x,y) = **0.0**

corr(x,y) = **0.0**

Euclidean (x,y) = **2.0**

d) x= (1,1,0,1,0,1) y=(1,1,1,0,0,1) cosine, correlation, Jaccard.

cos(x,y) = **0.75**

corr(x,y) = **0.25**

Jaccard (x,y) = **0.6**

e) x= (2,-1,0,2,0,-3) y= (-1,1,-1,0,0,-1) cosine, correlation.

cos(x,y) = **0.0**

corr(x,y) = -**0.0**

4. For the following vectors: x = (2,3,1,0,0), y = (0,1,4,2,0), z = (2,0,0,2,3), calculate:

a) The Euclidean distances d(x,y), d(y,z), d(x,z);

```
------------------
Euclidean distance
------------------
```

	x	y	z
x	0.000000	4.582576	4.795832
y	4.582576	0.000000	5.477226
z	4.795832	5.477226	0.000000

b) The cosine similarity cos(x, y) , cos(y, z) , cos (x, z);

cos(x,y) = **0.408248290463863**

cos(y,z) = **0.2117024496099853**

cos(x,z) = **0.2592814894208657**

c) The Jaccard coefficient of the binarized vectors $J(x_B, y_B), J(y_B, z_B), J(x_B, z_B)$, where the binarized vector x_B has entry 1 for each non-zero entry of x (and 0 for each zero entry of x).

$J(x_B, y_B)$ = **0.5**

$J(y_B, z_B) =$ **0.2**

$J(x_B, z_B) =$ **0.2**