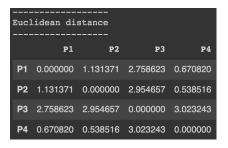
## **CI603 Data Mining**

## Tutorial 2 Part II

(solution)

 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.

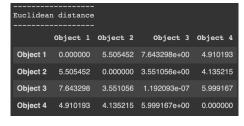
	X1(x)	X2(y)
P1	2.3	4
P2	1.5	3.2
P3	4.2	2
P4	1.7	3.7

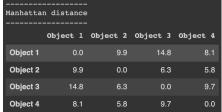




2. Using the table below calculate the Euclidean and Manhattan distance of the below data.

	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





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);

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).

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

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

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