THE CHINESE UNIVERSITY
OF HONG KONG (SHENZHEN)
School of Science and Engineering

Tutorial - 8

Fundamental of Artificial Intelligence **CSC3180**

1. **Answer the following questions:**

   (a) What is data mining? What is not data mining? Please give some examples. (L9: P4-5)

   (b) Why should we need to mine data? Please explain it in commercial and scientific viewpoints. (L9: P10-13)

   (c) How many types of data? Please give some examples about these types of data? (L9: P21-26)

   (d) We have learned about Minkowski distance on data measurement. What is the difference among L_1, L_2 and L_∞. (L9: P34)

   (e) Please describe the main steps of k-means algorithm. (L9: P47-48)

   (f) What about data exploration: regression? Please descript the simple linear regression.

2. **Please try to calculate the Minkowski Distance between each two variables when r = 1, 2, ∞.**

   a = [1, 0, 3, 1]; b = [1, 2, 5, 0]; c = [0, 4, 3, 1]

Solution:

L_1	a	b	c
a	0	5	5
b	5	0	6
c	5	6	0

L_2	a	b	c
a	0	3	4.12
b	3	0	3.16
c	4.12	3.16	0

L_∞	a	b	c
a	0	2	4
b	2	0	2
c	4	2	0

3. **Give the covariance matrix, please calculate the Mahalanobis Distance between each two inputs.**

$$\Sigma = \begin{bmatrix} 0.3, & 0.2 \\ 0.2, & 0.3 \end{bmatrix}$$

$a = [5,\ 4]$, $b = [6,\ 7]$, $c = [5,\ 6]$

Solution:

We can get the inverse of covariance matrix $\Sigma^{-1} = \begin{bmatrix} 6, & -4 \\ -4, & 6 \end{bmatrix}$. Then

$$D(a,b)=\sqrt{([5,4]-[6,7])*\begin{bmatrix}6, & -4\\ -4, & 6\end{bmatrix}*(\begin{bmatrix}5\\4\end{bmatrix}-\begin{bmatrix}6\\7\end{bmatrix})}=6$$

$$D(b,c)=\sqrt{([6,7]-[5,6])*\begin{bmatrix}6, & -4\\ -4, & 6\end{bmatrix}*(\begin{bmatrix}6\\7\end{bmatrix}-\begin{bmatrix}5\\6\end{bmatrix})}=2$$

$$D(a,c)=\sqrt{([5,4]-[5,6])*\begin{bmatrix}6, & -4\\ -4, & 6\end{bmatrix}*(\begin{bmatrix}5\\4\end{bmatrix}-\begin{bmatrix}5\\6\end{bmatrix})}=4.8990$$

4. **Assume X = [2 1 0 8 0 6 3 2 0 7] and Y = [3 3 4 2 0 1 3 0 0 5]. Please calculate their Cosine Similarity.**

Solution:

X.Y= 2*3 + 1*3 + 0*4 + 8*2 + 0*0 + 6*1 + 3*3 + 2*0 + 0*0 + 7*5 = 75

||X|| = (2*2+1*1+0*0+8*8+0*0+6*6+3*3+2*2+0*0+7*7)^0.5 = (167)^0.5 = 12.9228

||Y|| = (3*3+3*3+4*4+2*2+0*0+1*1+3*3+0*0+0*0+5*5)^0.5 = (73)^0.5 = 8.5440

Cosine(X,Y) = 75/(12.9228*8.5440) = 0.6793

5. **Assume that there are three mean vectors shown as follows. Give three inputs, please judge these points belong to which group using Euclidean Distance.**

Means: $\mu_1=[0,1,8,6,3,9]$ $\mu_2=[0,4,3,7,7,1]$ $\mu_3=[4,4,6,7,7,2]$

Inputs: $a_1=[6,6,1,1,4,9]$ $a_2=[3,5,2,7,2,3]$ $a_3=[6,8,9,5,1,3]$

Solution:

$D(a_1,\mu_1)=11.6619$, $D(a_1,\mu_2)=12.3693$, $D(a_1,\mu_3)=11.2694$

Since $D(a_1,\mu_3)<D(a_1,\mu_1)<D(a_1,\mu_2)$, a_1 belongs to the third group.

Similarly,

$D(a_2,\mu_1)=9.9499$, $D(a_2,\mu_2)=6.3246$, $D(a_2,\mu_3)=6.6332$

a_2 belongs to the second group.

$D(a_3,\mu_1)=11.2694$, $D(a_3,\mu_2)=11.4891$, $D(a_2,\mu_3)=8.3666$

a_3 belongs to the third group.