

## Proximity

1) similarity (s) → دوال يقبض على تآبيب two obj مع بعض

2) dissimilarity (d) دوال تبجب الاختلاف بين two obj مع بعض

في الحالتين بيطلع رقم بيتقوى شبه بعض او لا

بقاربه كل value من ال object الاول لا value  
التي في نفس ال attr في ال object الثاني

similarity  $\left\{ \begin{array}{l} 1 \text{ similar} \\ \vdots \\ 0 \text{ not similar} \end{array} \right.$   $\uparrow$  ميزه  
[0 - 1]

dissimilarity  $\left\{ \begin{array}{l} 0 \text{ similar} \\ (1 \text{ or not}) \text{ not similar} \end{array} \right.$   
(معايشه)  
upper limit

How to calculate similarity and dissimilarity

on nominal

$$S(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$$

where  $x$  is obj 1 value,  $y$  obj 2 value

$$d(x, y) = \begin{cases} 1 - S & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$$

$$d(x, y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$$

Example	Gender		Gender
obj 1	Male	-	obj 2
		-	Female

$$S(x, y) = 0$$

$$d(x, y) = 1$$



Calculate similarity or dissimilarity on ordinal values.

Example.

$A = \{S, M, L, XL\}$  int أجولها

$S \rightarrow 0$        $L \rightarrow 2$   
 $M \rightarrow 1$        $XL \rightarrow 3$

$$d(x, y) = \frac{|x - y|}{\text{range}} = \frac{|x - y|}{\text{Max} - \text{Min}} = \frac{|x - y|}{(n-1) - 0}$$

$$= \frac{|0 - 2|}{3} = \frac{2}{3}$$

$$s(x, y) = 1 - d(x, y) = \frac{1}{3}$$

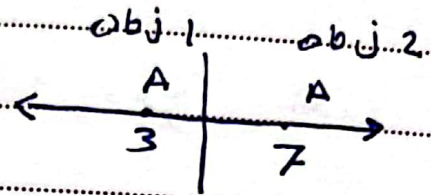
obj 1

obj 2

S	---
L	---

calculate similarity and dissimilarity on interval values.

$$d(x, y) = |x - y|$$



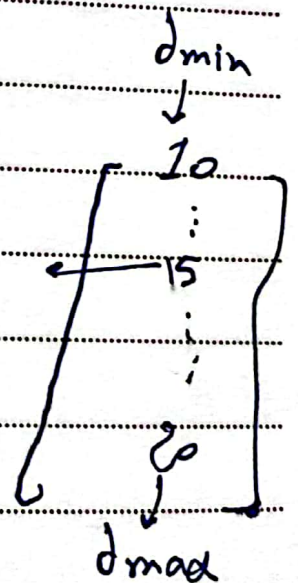
المسألة هنا انه في الاعداد فالبعض  
يوجد وال similarity لازم من [0 - 1]

Transformation في diss [0 - 1]

$$d_{new} = \frac{d_{old} - d_{min}}{d_{max} - d_{min}}$$

Example

$$d_n = \frac{15 - 10}{20 - 10} = \frac{5}{10} = 0.5$$





$$1 - \gamma = \frac{1}{1+d}$$

$$2 - \gamma = 1 - \frac{d_{old} - d_{min}}{d_{max} - d_{min}}$$

الاول يستخدمه لما من يعرف احد

Max او Min

الثانيه لما قيمته ال d تبقى اكبر  
من  $[0, 1]$  (و يبقى صدين ال Min و Max)

وعندي قانونه تالته  $\gamma = e^{-d}$

## Minkowski Distance

ال object ببص vector نظرا لأنه ببصوى  
على ارقام فقط

$$x = (x_1, x_2, x_3, x_4, \dots, x_n)$$

$$y = (y_1, y_2, \dots, y_n)$$

Note

\* clusters, ببصه ببص العلام

ال similar في مجموعته

\* classes, ببص العلام الى ببصهوا

في attr مبه في مجموعته

$$d(x, y) = \left( \sum_{i=1}^n |x_i - y_i|^r \right)^{1/r}$$

if  $r=1 \Rightarrow$  Manhattan  $\Rightarrow d(x, y) =$

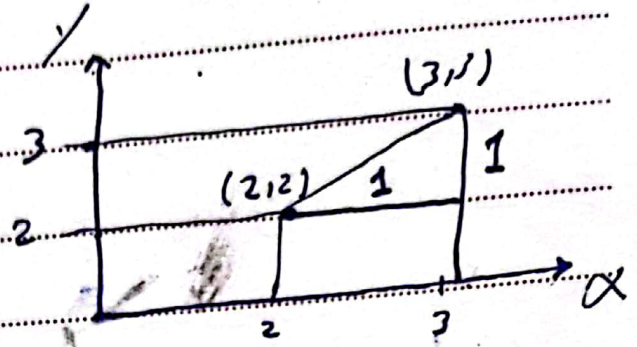
$$\sum_{i=1}^n |x_i - y_i|$$

Example -  $x = (1, 7, 8, 4)$ ,  $y = (5, 3, 2, 8)$

$$d(x, y) = |1-5| + |7-3| + |8-2| + |4-8| = 17$$



Example



Note - يفضل استخدام طريقة

$r=1$  الى فائته في حالة ال Binary

على وجه الخصوص (اقدم ما فته) بين نقطتين

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

Example 1

$$d(v_1, v_2)$$

$$\sqrt{(1-10)^2 + (5-20)^2 + (7-17)^2 + (2-13)^2 + (1-10)^2}$$

$$= 11$$

vec 1

vec 2

vec 3

	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$
vec 1	1	5	7	2	1
vec 2	10	20	17	13	10
vec 3	1	8	9	10	11

الرقم ده لو اغير

يبقوا similar لو كبر يبقوا

not similar

19014

almokhtabar.com

## \* Distance Matrix

$$\begin{matrix} & v_1 & v_2 & v_3 \\ v_1 & 0 & d(v_1, v_2) & d(v_1, v_3) \\ v_2 & d(v_2, v_1) & 0 & d(v_2, v_3) \\ v_3 & d(v_3, v_1) & d(v_3, v_2) & 0 \end{matrix}$$

المثلث الى فوق

هو الى تحته

و يتكرر بواحد بين

Triangle Matrix

بسيط Metric

## \* Mahalarobis Matrix,

$$d(x, y) = \sqrt{(x - y)^T \Sigma^{-1} (x - y)}$$

في حالة ال Binary Vector

Similarity

symmetric

SMC

asymmetric

Example,

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$f_{00} = 7, f_{01} = 2, f_{10} = 1, f_{11} = 0$$

$$SMC = \frac{f_{00} + f_{11}}{f_{01} + f_{10} + f_{00} + f_{11}} = \text{similarity}$$



$$d = 1 - S = \frac{f_{01} + f_{10}}{f_{00} + f_{11} + f_{01} + f_{10}}$$

بعبارة أخرى asymmetric لا يستخدم J

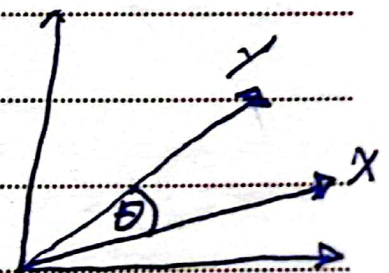
$$J = \frac{f_{11}}{f_{01} + f_{10} + f_{11}} \quad \left( \begin{array}{l} \text{نميلنا الى } f_{00} \text{ على حساب } f_{11} \\ \text{منه عايزين في asymmetric} \end{array} \right)$$

Note, SMC = simple matching coefficient

		$t_1$	$t_2$	$t_3$	$t_4$
	Doc 1	7	8	10	0
X	Doc 2	1	8	11	0
X	Doc 3	1	1	2	0

عنصر احصاف كثير وده مايدلش انهم شبيه  
بعض و بتعزم ال  $\cos(x, y)$

if  $\theta = 0$  ,  $\cos(\theta) = 1 \approx \text{Similarity (S)}$



$$\cos(x, y) = \frac{x \cdot y}{\|x\| \cdot \|y\|}$$

$x \cdot y = \text{inner Product } (x, y)$

~~B~~

$$x \cdot y = 1x_1 + 3x_1 + 0x_0 + 4x_2 = \underline{12}$$

$$x \cdot x = 26$$

$$\|x\| = \sqrt{26}$$

$\|y\| = \text{same as } x$

و نريد في قوسين

$v_1 \quad v_2 \quad v_3$

$$\begin{bmatrix} \cos(v_1, v_1) & \cos(v_1, v_2) & \cos(v_1, v_3) \\ \cos(v_2, v_1) & \cos(v_2, v_2) & \cos(v_2, v_3) \\ \cos(v_3, v_1) & \cos(v_3, v_2) & \cos(v_3, v_3) \end{bmatrix}$$