

(Similarity & Data Warehouse)

T. Nora

25-1-2024

محاضرة تعويضية

* Apriori Algorithms: 0

دراسة العلاقات بين العناصر

* Theory: $X \rightarrow Y$

التقاطع مفهوم

E	A	B	C	D
ball	apple	apple	new	apple
basket	relause	relause	apple	reCeipe
	new ipad	new iPad	receipe	new ipad

$$S(A, E) = 0$$

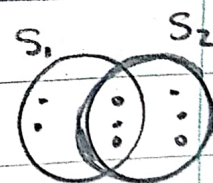
A with all is Zero

$$S(A, B) = 3$$

$$S(A, D) = 2$$

$$S(B, C) = 2$$

$A = \{Apple, new iPad, \dots\}$



$$J_{sim}(S_1, S_2) = J_{sim}(S_2, S_1)$$

* between vector \Rightarrow

نسبته بين مجاميع

1-d
x

2d
y L x

3d
y
z x

الواجب:

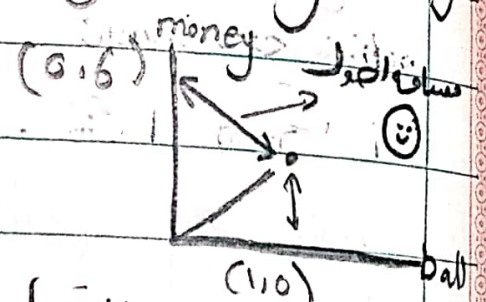
	apple	orange	banana	Mango	
d1	10	30	0	0	1
d2	20	10	10	0	2
d3	0	0	0	20	7

Similarity \Rightarrow d1 and d2

* B "money, money, money" money, money, money

A "ball"

C "money money ball"



• Similarity \Rightarrow $\frac{1}{\sqrt{2}}$

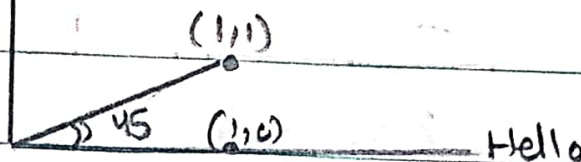
Cos \rightarrow $0 \rightarrow 1$

* A "Hello world"

B "Hello"

world

	Hello	world
A	1	1
B	1	0



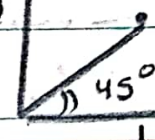
$$\cos(45^\circ) = \frac{1}{\sqrt{2}} = 0.7$$

* A "Hello world"

B "Hello Hello Hello"

	H	W
A	1	1
B	3	0

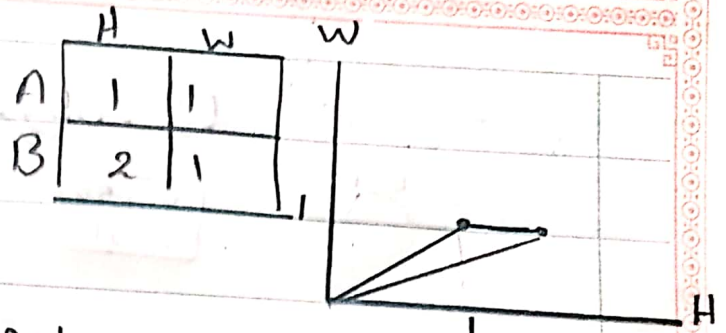
السكرانة لثلاثون على لثلاثون



الواجب:

* A "Hello World"

B "Hello Hello World"



* معيار الموهبة في هذه الخوارزمية = 4
 الجرد @ مقدار الزاوية θ
 Cos θ

* A "Hello World"

A(1,1)

B "Hello Hello Hello"

B(3,0)

$$\sum_{i=1}^n A_i B_i$$

$$\sqrt{\sum_{i=1}^n A_i^2} * \sqrt{\sum_{i=1}^n B_i^2}$$

A_i

index

قائمة

$$* (1*3) + (1*0)$$

$$\sqrt{1^2 + 1^2} * \sqrt{3^2 + 0^2}$$

$$= \frac{3}{\sqrt{2} \sqrt{9}} = \frac{3}{3\sqrt{2}} = \frac{1}{\sqrt{2}}$$

* "I love Trello 2"

"I love Gys"

	I	love	Trello	2	Gys
A	1	1	1	1	0
B	1	1	0	0	1

$$(1*1) + (1*1) + (1*0) + (1*0) + (0*1)$$

$$= 6.58$$

$$\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 0^2} * \sqrt{1^2 + 1^2 + 0^2 + 0^2 + 1^2}$$

الواجب:

* فتحت مواقع تم إرسالها للقربان .

1800 مستخدمين

0, 90
0 → 1

* Similarity ⇒ unsupervise

* AI - Face Reco ⇒ similarity نستخرج

* Data Base ⇒ insert - update - Delete - select ⇒ optional

* Data Warehouse ⇒ insert - select ⇒ historical Data

* Data Warehouse Schema ⇒ Normalization

A	1	1	منار
B	1	2	خولة
C	2	0 → 3	

تقارير

انواع الاجازة

* Data warehouse

* largest data

* Star

* Snowflake

اسم الموظف	مدير	عمره	حج

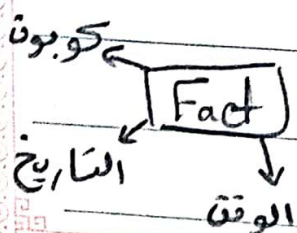
* target Data ⇒

كل حاجة نطرحها بعض

اجازة مرضية - اجازة حج - اجازة طاريء .

* dimension ⇒ حرة - مدينة - وقت العمل

نوع الإدخال - نوع الكوبون



الواجب:

كل حاجة نحدد هان نقول الرسل
 1) كم بعنا بروتة الصباغ \Rightarrow نوع البركة
 2) كم عدد الكوبونات \Rightarrow اي يوتيوب مثالي \Rightarrow نوع الكوبون

* نظام مبيعات \Rightarrow بيع online \Rightarrow تساهيل مثالي

* BI \Rightarrow Beta Base

tools فقد نندش question

منار قاضي
 CS4