Lecture-4 (Part-2) Math

Apriori Example

Ex: 6.3 (Table-6.1)

Data Mining Book (P-250)

1st Scan

C_1

Iteam Set	Sup. count
$\{I_1\}$	6
$\{I_2\}$	7
$\{I_3\}$	6
$\{I_4\}$	2
$\{I_5\}$	2

→ compare candidate support count with min

L_1

Iteam set	Sup. count
$\{I_1\}$	6
$\{I_2\}$	7
$\{I_3\}$	6
$\{I_4\}$	2
$\{I_5\}$	2

2nd Scan

C_2 (Generate "C_2" candidate from L_1)

Iteam Set	Sup. count
$\{I_1, I_2\}$	4
$\{I_1, I_3\}$	4
$\{I_1, I_4\}$	1 ✗
$\{I_1, I_5\}$	2
$\{I_2, I_3\}$	4
$\{I_2, I_4\}$	2
$\{I_2, I_5\}$	2
$\{I_3, I_4\}$	0 ✗
$\{I_3, I_5\}$	1 ✗
$\{I_4, I_5\}$	0 ✗

→ compare candidate support count with min

L_2

Iteam Set	Sup. count
$\{I_1, I_2\}$	4 ✓
$\{I_1, I_3\}$	4 ✓
$\{I_1, I_5\}$	2
$\{I_2, I_3\}$	4 ✓
$\{I_2, I_4\}$	2
$\{I_2, I_5\}$	2

Generate "C_3" candidate from "L_2"

C_3

Iteam set	Sup. count
W I_1, I_2, I_3	2
I_1, I_2, I_4	1
W I_1, I_2, I_5	2
✗ I_1, I_3, I_4	0
✗ I_1, I_3, I_5	1
✗ I_1, I_4, I_5	0
✗ I_2, I_3, I_4	0
✗ I_2, I_3, I_5	1
✗ I_2, I_4, I_5	0
I_3, I_4, I_5	0

compare
Candidate
Suppont
Count with min

→

L_3

Iteam set	Sup. count
I_1, I_2, I_3	2
I_1, I_2, I_5	2

C_4

Iteam set	Sup. count
I_1, I_2, I_3, I_4	0
I_1, I_2, I_3, I_5	1
I_1, I_2, I_4, I_5	0

Here, we can see th. for the C_4 iteam set support count is below the required sup. count. doesn't satisfy the minimum support. So, we can't take it as an iteam set.
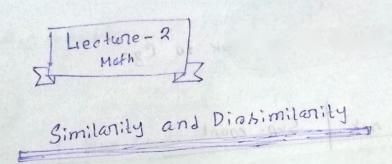
Now, we need to go back to C_3.

Iteam set	Sup. count
I_1, I_2, I_3	2
I_1, I_2, I_5	2

So, frequent-3 ~~(C_3)~~ iteam set, $I = \{I_1, I_2, I_3\}$ and $\{I_1, I_2, I_5\}$

Minimum Support $= 2/9 = 22.22\%$

$\not A b_2$

$\{I_1, I_2, I_3\}$.

$\quad \longrightarrow \{I_1; I_2\} \rightarrow I_3$

$\quad\quad \{I_1, I_3\} \rightarrow I_2$

$\quad\quad \{I_2, I_3\} \rightarrow I_1$

$\quad\quad I_1 \rightarrow \{I_2, I_3\}$

$\quad\quad I_2 \rightarrow \{I_1, I_3\}$

$\quad\quad I_3 \rightarrow \{I_1, I_2\}$

$\{I_1, I_2, I_5\}$

$\quad \longrightarrow \{I_1, I_2\} \rightarrow I_5$

$\quad\quad \{I_1, I_5\} \rightarrow I_2$

$\quad\quad \{I_2, I_5\} \rightarrow I_1$

$\quad\quad I_1 \rightarrow \{I_2, I_5\}$

$\quad\quad I_2 \rightarrow \{I_1, I_5\}$

$\quad\quad I_5 \rightarrow \{I_1, I_3\}$

Am:

Similarity and Dissimilarity

Ex: 2·18 (Table-2·3)

(P-70)

Sol^n: Let, $q(1,1) = 1$

$n(1,0) = 1$

$s(0,1) = 1$

$t(0,0) = 1$

Dissimilarity between i and j:

$$d(i,j) = \frac{n+s}{q+n+s+t} \quad \left(\begin{array}{l}\text{symmetric binary}\\ \text{dissimilarity}\end{array}\right)$$

$$= \frac{1+1}{1+1+1+1}$$

$$d(i,j) = \frac{n+s}{q+n+s} \quad \left(\begin{array}{l}\text{\$Asymmetric binary}\\ \text{similarity})\end{array}\right)$$

$$sim_{Jaccard}(i,j) = \frac{q}{q+n+s} \quad \left(\begin{array}{l}\text{\$ Jaccard coefficient}\\ \text{(similarity measure}\\ \text{for asymmetric}\\ \text{binary values)}\end{array}\right)$$

know it

Symmetric	Asymmetric
If both of the ~~N status~~ on values are equally important.	If both of the status on values are not equally important.
//(We can loss information if we exclude Value).	Ex: positive and negative outcome of a disease.
Ex: Gender: Male Female	//(we may not loss information if we exclude any value).

Table - 2.4 W

§ Soln: From the given table we found that -

◉ Gender is symmetric attribute.

◉ The remaining attributes are asymmetric binary.

Let, Y = 1 and N = 0
P = 1

$q(1,1) = 1$, $r(1,0) = 1$, $S(0,1) = 1$, $t(0,0) = 1$

d̂(Jack, Mary) ̃~~~~~~~

Now, finding dissimilarity using asymmetric attributes.

$$d(\text{jack, mary}) = \frac{0+1}{2 \cdot 0 + 0 + 1} = \frac{1}{3} = 0.33$$

$$d(\text{jack, jim}) = \frac{1+1}{1+1+1} = \frac{2}{3} = 0.67$$

$$d(\text{jim, mary}) = \frac{1+2}{1+1+2} = \frac{3}{4} = 0.75 \qquad \underline{Am:}$$

Cosine Similarity

cosine measure: If d_1 and d_2 are two vectors

(e.g., term-frequency vectors), then

$$\cos(d_1, d_2) = (d_1 \bullet d_2) / \|d_1\| \, \|d_2\|,$$

Where "$\bullet$" : Indicates vector dot product.

$\|d\|$: Length of vector d.

Ex: Similarity between documents 1 and 2.

$$d_1 = (5, 0, 3, 0, 2, 0, 0, 2, 0, 0)$$
$$d_2 = (3, 0, 2, 0, 1, 1, 0, 1, 0, 1)$$

$d_1 \bullet d_2 = 5*3 + 0*0 + 3*2 + 0*0 + 2*1 + 0*1 +$
$\qquad 0*0 + 2*1 + 0*0 + 0*1 = 25$

$\|d_1\| = 5*5 + 0*0 + 3*3 + 0*0 + 2*2 + 0*0 + 0*0 + 2*2 + 0*0 + 0*0$
$\qquad = (42)^{0.5} = 6.48$

$\|d_2\| = 3*3 + 0*0 + 2*2 + 0*0 + 1*1 + 1*1 + 0*0 + 1*1 + 0*0 + 1*1$
$\qquad = (17)^{0.5} = 4.12$

$\cos(d_1, d_2) = 25/(6.48*4.12) = 0.94 \qquad \underline{Am:}$