DMBI

Meghpara Jeet J.
160410116056

* ASSIGNMENT-5 *

Date: 14 IT-1

(1) Find frequent item sets using apriori for follo market basket analysis data. Let minimum thresold is 3. At each step show first apriori based pruning & then form freq itemset.

TID	Items_bought
T_1	{1, 2, 3, 4, 5, 6}
T_2	{7, 2, 3, 4, 5, 6}
T_3	{1, 8, 4, 5}
T_4	{1, 9, 0, 4, 6}
T_5	{0, 2, 2, 4, 5}

→ Step 1. (1-Itemset)

	Count			
I_1	3		I_1	3
I_2	4		I_2	4
I_3	2		I_3	2
I_4	5		I_4	5
I_5	4	→	I_5	4
I_6	3		I_6	3
I_7	1		I_7	1
I_8	1		I_8	1
I_9	1		I_9	1
I_{10}	2		I_{10}	2

YASH

Royal ECO

Date:

Step 2. (2 itemset freq pattern)

	Count
{I1, I2}	1
{I2, I3}	2
{I3, I4}	2
{I2, I4}	3
{I2, I5}	3
{I2, I6}	2
{I3, I5}	2
{I3, I6}	2
{I4, I6}	3
{I4, I5}	4
{I5, I6}	2
{I1, I4}	3
{I1, I6}	2
{I1, I5}	2

→

	Count
{I2, I4}	3
{I2, I5}	3
{I4, I6}	3
{I4, I5}	4
{I1, I4}	3

Step 3. (3 itemset freq pattern)

	Count
{2, 3, 4}	2
{2, 3, 5}	2
{2, 3, 6}	2
{3, 4, 5}	2
{3, 4, 6}	2
{4, 5, 6}	2
{1, 4, 5}	2
{2, 5, 6}	2
{2, 4, 6}	2
{2, 4, 5}	3

→

	Count
{2, 4, 5}	3
{1, 4, 5}	2
{4, 5, 6}	2

as minimum threshold is 3.

✓

	Count
{2, 4, 5}	3

Date:

So finally,
we have $C_2 = \{\{2,4,5\}\}$

(2) Use two methods below to normalize following group of data.
200, 300, 400, 600, 1000.

(a) min-max by setting min = 0 & max = 1

→ $$v' = \frac{v - min_A}{max_A - min_A}\left(new(max_A - min_A)\right) + new\,min_A$$

$v_1 = 200$ $min_A = 200$
$v_2 = 300$ $max_A = 1000$
$v_3 = 400$
$v_4 = 600$ $new\,min_A = 0$
$v_5 = 1000$ $new\,max_A = 1$

For v_1,
$$v_1' = \frac{(200-200)}{1000-200}(1-0) + 0$$
$$= 0.$$

For v_2,
$$v_2' = \frac{(300-200)}{1000-200}(1-0) + 0.$$
$$= 1/8.$$

For v_3,
$$v_3' = \frac{(400-200)}{(1000-200)}(1-0) + 0$$
$$= 2/8 = 1/4$$

YASH

Royal ECO

Date:

For V_4,

$$V_4' = \frac{600-200}{1000-200}(1-0)+0$$

$$= 4/8 = 1/2$$

For V_5,

$$V_5' = \frac{1000-200}{1000-200}(1-0)+0$$

$$= 1$$

(b) z-score normalization.

→ $$\text{z-score} = \frac{x-\mu}{\sigma}$$

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$$\mu = \frac{200+300+400+600+1000}{5}$$

$$= 500$$

$$\sigma = \sqrt{\frac{(500-200)^2+(200)^2+(100)^2+(100)^2+(500)^2}{5}}$$

$$= \sqrt{\frac{90000+40000+10000+10000+250000}{5}}$$

$$= 282.84.$$

Date:

for x_1, $\frac{200 - 500}{282.84}$

$x_1 = -1.06$.

for x_2,

$x_2 = \frac{300 - 500}{282.84}$

$x_2 = -0.707$

for x_3, $\frac{400 - 500}{282.84}$

$x_3 = -0.3535$

for x_4, $\frac{600 - 500}{282.84}$

$x_4 = 0.3535$.

for x_5,

$= \frac{1000 - 500}{282.84}$

$x_5 = 1.7677$.