* Consider a set of x'actions given :

TID	items
T_1	i_1 i_2 i_5
T_2	i_2 i_4
T_3	i_2 i_3
T_4	i_1 i_2 i_4
T_5	i_1 i_3
T_6	i_2 i_3
T_7	i_1 i_3
T_8	i_1 i_2 i_3 i_5
T_9	i_1 i_2 i_3

→ Let minimum support = 2.

[If minimum support is given in % say 30% then it is actually support threshold & minimum support can be found by product of support threshold & no. of x'actions. In this example, it is directly given as 2.]

step I scan the db & find set of candidate 1-item set, C1, & their support counts.

$\therefore$ C1 = { {i_1}, {i_2}, {i_3} {i_4}, {i_5}}

S.C. = 6 7 6 2 2

[Each individual item is member of C1]

Now, compare their support count with minimum support so as to discard those items that are not frequent & to find set of frequent 1-item sets, L_1

In this example, they all have support count >= min. support (2). That is, they all are frequent

$\therefore L_1 = C_1 = \{\{i_1\}, \{i_2\}, \{i_3\}, \{i_4\}, \{i_5\}\}$

<u>step II</u> Now to find C_2, set of candidate 2-item sets, join L_1 with itself.

$C_2 = L_1 \bowtie L_1$

$= \{\{i_1, i_2\}, \{i_1, i_3\}, \{i_1, i_4\} \{i_1, i_5\}$
$\{i_2\ i_3\}, \{i_2\ i_4\} \{i_2\ i_5\}$
$\{i_3\ i_4\}, \{i_3\ i_5\}$
$\{i_4\ i_5\}\}$

[Basically, you have to combine each item, with ~~so~~ rest of each items]

Now scan the db to find their support counts

	sc		
$\therefore C_2 = \{\{i_1\ i_2\} \rightarrow$	4	$\{i_2\ i_3\} \rightarrow 4$	$\{i_3\ i_5\} \rightarrow 1$
$\{i_1\ i_3\} \rightarrow$	4	$\{i_2\ i_4\} \rightarrow 2$	$\{i_4, i_5\} \rightarrow 0$
$\{i_1\ i_4\} \rightarrow$	1	$\{i_2\ i_5\} \rightarrow 2$	$\}$
$\{i_1\ i_5\} \rightarrow$	2	$\{i_3\ i_4\} \rightarrow 0$	

Now, discard those with support count less than min_support to get L_2

$\therefore L_2 = \{ \{i_1, i_2\} \rightarrow 4$

$\{i_1, i_3\} \rightarrow 4$

$\{i_1, i_5\} \rightarrow 2$

$\{i_2\ i_3\} \rightarrow 4$

$\{i_2\ i_4\} \rightarrow 2$

$\{i_2\ i_5\} \rightarrow 2$

$\}$

<u>step III</u> Now join L_2 with L_2 to get C_3

$\therefore C_3 = L_2 \bowtie L_2$

[Now, while joining L_{k-1} with L_{k-1}, only those members are joinable whose first $k-2$ elements are common]

In, this $k = 3$ as we are generating C_3

$\therefore k-2 = 1$, that is, when we join L_2 with L_2, only those members can be joined whose first one element ($\because k-2=1$) is common. for ex $\{i_1, i_2\}$ & $\{i_1, i_3\}$ are joinable as first one element i_1 is common. But $\{i_1, i_2\}$ & $\{i_2\ i_3\}$ are not joinable as their first one element is not common.

This condition is required so as to avoid generating duplicate elements in C_k.

$\therefore C_3 = L_2 \bowtie L_2$

$= \{ \{i_1\ i_2\ i_3\}$ 1^{st} & 2^{nd} of L_2

$\{i_1\ i_2\ i_5\}$ 1^{st} & 3^{rd} of L_2

$\{i_1\ i_3\ i_5\}$ 2^{nd} & 3^{rd} of L_2

$\{i_2\ i_3\ i_4\}$ 4^{th} & 5^{th} of L_2

$\{i_2\ i_3\ i_5\}$ 4^{th} & 6^{th} of L_2

$\{i_2\ i_4\ i_5\}$ 5^{th} & 6^{th} of L_2

$\}$

Now use apriori ~~prop~~ property before you decide to scan the db for each member of C_3.

	SC	
$\therefore C_3 = \{ \{i_1\ i_2\ i_3\}$	2	[DB scan is reqd. as all non-empty subsets are frequent]
$\{i_1\ i_2\ i_5\}$	2	[——— " ———]
$\{i_1\ i_3\ i_5\}$	N.A.	[It is discarded by apriori property, as one of its subsets $\{i_3\ i_5\}$ is not frequent, no. db scan is performed]

SC

{i_2 i_3 i_4} N.A. [—— " ——]

{i_2 i_3 i_5} N. A. [—— " ——]

{i_2 i_4 i_5} N.A. [—— " ——]

}

∴ We are left with only 2 members in C_3

$$\therefore C_3 = \{ \{i_1 \; i_2 \; i_3\} \rightarrow 2 \quad \{i_1 \; i_2 \; i_5\} \rightarrow 2 \}$$

(SC)

Now, compare their support count with min support, so as to discard those which are not frequent.

∴ they both have support count >= min support. So, they both are frequent

$$\therefore L_3 = \{ \{i_1 \; i_2 \; i_3\} \rightarrow 2 \quad \{i_1 \; i_2 \; i_5\} \rightarrow 2 \}$$

step IV Now $C_4 = L_3 \bowtie L_3$

here $k=4$ ∴ $k-2=2$

∴ only those members of L_3 are joinable whose first 2 elements are common (as $k-2=2$).

$\{i_1\ i_2\ i_3\}$ & $\{i_1\ i_2\ i_5\}$ are joinable as their first elements i_1, i_2 are common.

$\therefore C_4 = L_3 \bowtie L_3$

$= \{ \{i_1\ i_2\ i_3\ i_5\}$

$\}$

Apply apriori property to check if this item set can be discarded directly or db scan is required

One of it's non-empty subsets $\{i_3\ i_5\}$ is not frequent. So, this itemset can not be frequent

It is discarded directly without scanning the db

$\therefore L_4 = \phi$.

Algo terminates, as now no further ~~?~~ C_k or L_k can be generated.

Scanned by CamScanner