Homework 5

Due: 05/09/2022

Devika Chandnani
A13405666

Q1

a) The data given has 6 features : $I_1 - I_6$

	I_1	I_2	I_3	I_4	I_5	I_6
1	0	0	1	0	1	0
2	0	1	1	1	0	1
3	1	0	0	0	1	0
4	1	1	1	0	0	0
5	0	0	0	1	0	0
6	1	0	0	1	0	1
7	0	0	1	1	1	1
8	1	0	1	0	1	0
9	1	0	0	1	0	0
10	0	1	1	0	0	1

* Apriori algorithm follows the logic that an itemset is not frequent if any of it's subsets are infrequent

* K = 1

Item Set	Support count
I_1	5
I_2	3
I_3	6
I_4	5
I_5	4
I_6	4

* In this, all items hold the minimum treshold of three transactions

* K = 2

Itemset	Support count
$\{I_2, I_3\}$	3
$\{I_3, I_5\}$	3
$\{I_3, I_6\}$	3
$\{I_4, I_6\}$	3

* From ${}^6C_2$ pairs of two, the ones above hold the minimum treshold count of 3

* K = 3

Itemset	Support count
$\{I_3, I_5, I_6\}$	1

* This item is below the minimum treshold count so this is an empty table

→ Therefore, we got 4 maximal frequent sets in level 2 when K=2 whose maximal support count is 3.

b) Pick one maximal set and check if it's subsets are association rules with frequency ≥ 0.3 and confidence ≥ 0.6

1) I_2, I_3

$$\text{confidence } I_2 \rightarrow I_3 = \frac{\text{support count } (I_2 \cup I_3)}{\text{support count } (I_2)} \times 100$$
$$= 3/3 \times 100 = 100\%.$$
$$\text{confidence } I_3 \rightarrow I_2 = \frac{\text{support count } (I_2 \cup I_3)}{\text{support count } (I_3)} \times 100$$
$$= 3/6 \times 100 = 50\%. \quad \times$$

2) I_3, I_5

$$\text{confidence } I_3 \rightarrow I_5 = \frac{\text{support count } (I_3 \cup I_5)}{\text{support count } (I_3)} \times 100$$
$$= 3/6 \times 100 = 50\%. \quad \times$$
$$\text{confidence } I_5 \rightarrow I_3 = \frac{\text{support count } (I_3 \cup I_5)}{\text{support count } (I_5)} \times 100$$
$$= 3/4 \times 100 = 75\%.$$

3) I_3, I_6

$$\text{confidence } I_3 \rightarrow I_6 = \frac{\text{support count } (I_3 \cup I_6)}{\text{support count } (I_3)} \times 100$$
$$= 3/6 \times 100 = 50\%. \quad \times$$
$$\text{confidence } I_6 \rightarrow I_3 = \frac{\text{support count } (I_3 \cup I_6)}{\text{support count } (I_6)} \times 100$$
$$= 3/4 \times 100 = 75\%.$$

4) $I_4 \rightarrow I_6$

$$\text{confidence } I_4 \rightarrow I_6 = \frac{\text{support count } (I_4 \cup I_6)}{\text{support count } (I_4)} \times 100$$
$$= 3/5 \times 100 = 60\%.$$
$$\text{confidence } I_6 \rightarrow I_4 = \frac{\text{support count } (I_4 \cup I_6)}{\text{support count } (I_6)} \times 100$$
$$= 3/4 \times 100 = 75\%.$$

Q2 minimum support = 2

TID	Items
1	{a,b,e}
2	{a,b,c,d}
3	{a,c,d}
4	{a,c,e}
5	{b,c,f}
6	{a}
7	{a,b,c}
8	{b,d,e}
9	{a,c}
10	{a,b,d,e}

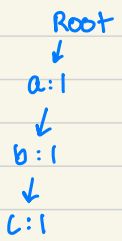
a) construct FP tree

Items	Support
a	8
b	6
c	6
d	4
e	4
f	①

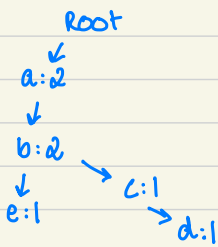
↳ less than min support !

* we remove 'f' from the database before constructing the tree

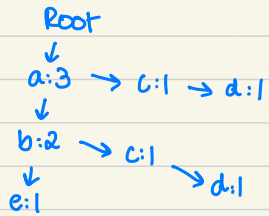
1) {a, b, e}



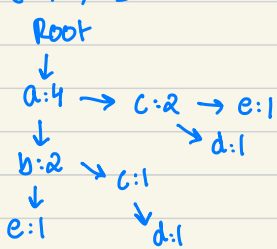
2) {a, b, c, d}



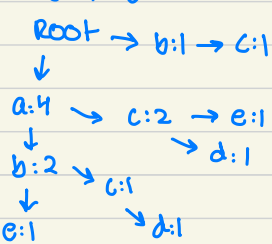
3) {a, c, d}



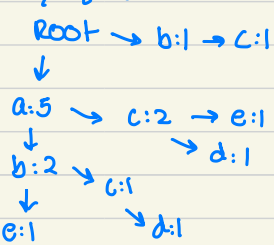
4) {a, c, e}



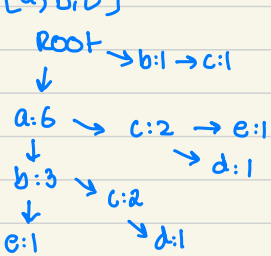
5) {b, c}



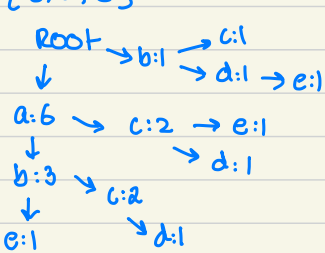
6) {a}



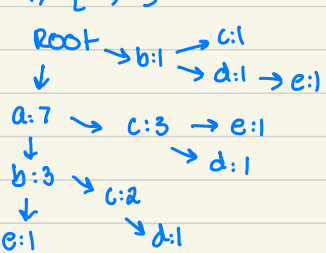
7) {a, b, c}



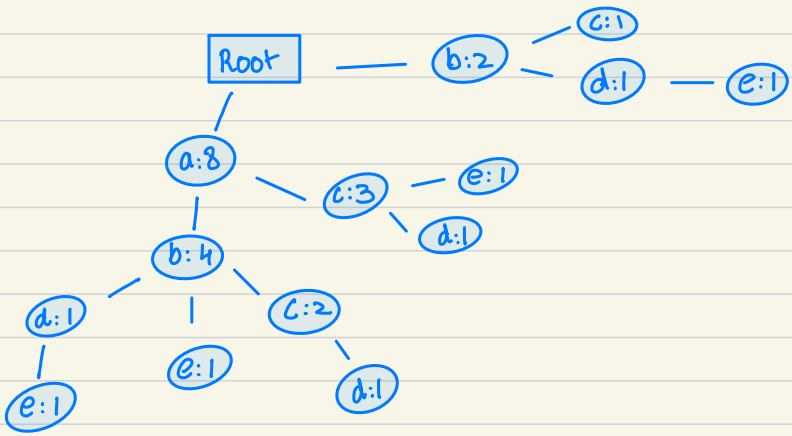
8) {b, d, e}



9) {a, c}



10) Final Tree {a, b, d, e}



b) 1) d's conditional pattern base : {a, b : 1} {a, b, c : 1} {a, c : 1} {b : 1}
This can be found by taking note of all items which are above d along with d's support count in that branch

2) d's conditional FP tree : {a : 3, b : 2, c : 2}

3) d's frequent patterns : {a, d : 3}
{b, d : 2}
{c, d : 2}
{a, b, c, d : 2}
{a, b, d : 2}
{b, c, d : 2}

Q3

L3 =
<{2}{3}{4}>
<{2 5}{3}>
<{3}{4}{5}>
<{1}{2}{3}>
<{1}{2 5}>
<{1}{5}{3}>
<{5}{3 4}>

<{1}{2}{3}>:
<{2}{3}{4}> ✓
<{2 5}{3}>
<{3}{4}{5}>
<{1}{2}{3}>
<{1}{2 5}>
<{1}{5}{3}>
<{5}{3 4}>

<{2}{3}{4}>:
<{2}{3}{4}>
<{2 5}{3}>
<{3}{4}{5}> ✓
<{1}{2}{3}>
<{1}{2 5}>
<{1}{5}{3}>
<{5}{3 4}>

<{1}{2 5}>:
<{2}{3}{4}>
<{2 5}{3}> ✓
<{3}{4}{5}>
<{1}{2}{3}>
<{1}{2 5}>
<{1}{5}{3}>
<{5}{3 4}>

<{2 5}{3}>:
<{2}{3}{4}>
<{2 5}{3}>
<{3}{4}{5}>
<{1}{2}{3}>
<{1}{2 5}>
<{1}{5}{3}>
<{5}{3 4}> ✓

<{1}{5}{3}>:
<{2}{3}{4}>
<{2 5}{3}>
<{3}{4}{5}>
<{1}{2}{3}>
<{1}{2 5}>
<{1}{5}{3}>
<{5}{3 4}> ✓

<{3}{4}{5}>:
<{2}{3}{4}>
<{2 5}{3}>
<{3}{4}{5}>
<{1}{2}{3}>
<{1}{2 5}>
<{1}{5}{3}>
<{5}{3 4}>

<{5}{3 4}>:
<{2}{3}{4}>
<{2 5}{3}>
<{3}{4}{5}>
<{1}{2}{3}>
<{1}{2 5}>
<{1}{5}{3}>
<{5}{3 4}>

Therefore, length 4 candidates L_4 = $<\{2\}\{3\}\{4\}\{5\}>$
$<\{2\,5\}\{3\,4\}>$
$<\{1\}\{2\}\{3\}\{4\}>$
$<\{1\}\{2\,5\}\{3\}>$
$<\{1\}\{5\}\{3\,4\}>$

Remove 1:

$<\{2\}\{3\}\{4\}\{5\}>$:
- ~~$<\{2\}\{3\}\{4\}\{5\}>$~~ → $<\{3\}\{4\}\{5\}>$ ✓
- $<\{2\}$ ~~$\{3\}$~~ $\{4\}\{5\}>$ → $<\{2\}\{4\}\{5\}>$
- $<\{2\}\{3\}$ ~~$\{4\}$~~ $\{5\}>$ → $<\{2\}\{3\}\{5\}>$
- $<\{2\}\{3\}\{4\}$ ~~$\{5\}$~~ $>$ → $<\{2\}\{3\}\{4\}>$ ✓

$<\{2\,5\}\{3\,4\}>$:
- ~~$<\{2\,5\}\{3\,4\}>$~~ → $<\{5\}\{3\,4\}>$ ✓
- $<\{2\,5\}$ ~~$\{3\,4\}$~~ $>$ → $<\{2\}\{3\,4\}>$
- $<\{2\,5\}$ ~~$\{3\}$~~ $\{4\}>$ → $<\{2\,5\}\{4\}>$
- $<\{2\,5\}\{3\}$ ~~$\{4\}$~~ $>$ → $<\{2\,5\}\{3\}>$ ✓

$<\{1\}\{2\}\{3\}\{4\}>$:
- ~~$<\{1\}\{2\}\{3\}\{4\}>$~~ → $<\{2\}\{3\}\{4\}>$ ✓
- $<\{1\}$ ~~$\{2\}$~~ $\{3\}\{4\}>$ → $<\{1\}\{3\}\{4\}>$
- $<\{1\}\{2\}$ ~~$\{3\}$~~ $\{4\}>$ → $<\{1\}\{2\}\{4\}>$
- $<\{1\}\{2\}\{3\}$ ~~$\{4\}$~~ $>$ → $<\{1\}\{2\}\{3\}>$ ✓

$<\{1\}\{2\,5\}\{3\}>$:
- ~~$<\{1\}\{2\,5\}\{3\}>$~~ → $<\{2\,5\}\{3\}>$ ✓
- $<\{1\}$ ~~$\{2\,5\}$~~ $\{3\}>$ → $<\{1\}\{5\}\{3\}>$ ✓
- $<\{1\}\{2\,5\}$ ~~$\{3\}$~~ $>$ → $<\{1\}\{2\}\{3\}>$ ✓
- $<\{1\}\{2\,5\}\{3\}>$ → $<\{1\}\{2\,5\}>$ ✓

$<\{1\}\{5\}\{3\,4\}>$:
- ~~$<\{1\}\{5\}\{3\,4\}>$~~ → $<\{5\}\{3\,4\}>$ ✓
- $<\{1\}$ ~~$\{5\}$~~ $\{3\,4\}>$ → $<\{1\}\{3\,4\}>$
- $<\{1\}\{5\}$ ~~$\{3\,4\}$~~ $>$ → $<\{1\}\{5\}\{4\}>$
- $<\{1\}\{5\}\{3\}$ ~~$\{4\}$~~ $>$ → $<\{1\}\{5\}\{3\}>$ ✓

↓ Check to see if the following have support from L_3

L_3 =
- $<\{2\}\{3\}\{4\}>$
- $<\{2\,5\}\{3\}>$
- $<\{3\}\{4\}\{5\}>$
- $<\{1\}\{2\}\{3\}>$
- $<\{1\}\{2\,5\}>$
- $<\{1\}\{5\}\{3\}>$
- $<\{5\}\{3\,4\}>$

∴ Therefore L_4 = $<\{1\}\{2\,5\}\{3\}>$

Q4 Given the time series:

$$X = [39\ 44\ 43\ 39\ 46\ 38\ 39\ 43]$$
$$Y = [37\ 44\ 41\ 44\ 39\ 39\ 39\ 40]$$

local cost function is given by: $c(x_i, y_i) = d(x_i, y_i) = |x_i, y_i|$

example: $d(39, 37) = 2$

* Calculate DTW distance between x and y and optimal warping path

Time series X

43	35	22	17	16	12	12	12	11
39	29	21	15	18	8	8	8	9
38	27	16	13	13	8	9	10	12
46	26	10	10	7	12	12	12	11
39	17	8	5	9	5	5	5	6
43	13	3	4	5	9	13	17	20
44	9	2	5	5	10	15	19	18
39	2	7	9	14	14	14	14	13
	37	44	41	44	39	39	39	40

→ optimal warping path

Time Series Y

∴ The DTW distance between X & Y = 11