

Computer Science 145, Homework 5

Michael Wu
UID: 404751542

March 6th, 2019

Problem 1

a) In the first iteration we obtain the following table.

Item Set	Support
$\{a\}$	6
$\{b\}$	8
$\{c\}$	6
$\{d\}$	4
$\{e\}$	2
$\{i\}$	1
$\{j\}$	1
$\{k\}$	1

We can ignore $\{i\}$, $\{j\}$, and $\{k\}$ since they are not frequent. Then we have the following length two candidates and supports.

Item Set	Support
$\{a, b\}$	4
$\{a, c\}$	4
$\{a, d\}$	2
$\{a, e\}$	2
$\{b, c\}$	4
$\{b, d\}$	4
$\{b, e\}$	2
$\{c, d\}$	1
$\{c, e\}$	1
$\{d, e\}$	0

We can ignore $\{c, d\}$, $\{c, e\}$, and $\{d, e\}$ since they are not frequent. Then we generate the length three candidates by looking at the length two frequent patterns and combining the ones that only differ in the last element. During candidate generation we can prune $\{a, c, d\}$, $\{a, c, e\}$, $\{a, d, e\}$, $\{b, c, d\}$, $\{b, c, e\}$, and $\{b, d, e\}$ since they are supersets of the previous length two sets that are not frequent. We then have the following length three candidates and supports.

Item Set	Support
$\{a, b, c\}$	2
$\{a, b, d\}$	2
$\{a, b, e\}$	2

During generation of length four candidates, we will prune away all the candidates because they contain one of $\{c, d\}$, $\{c, e\}$, or $\{d, e\}$ which are not frequent. Thus our frequent patterns are

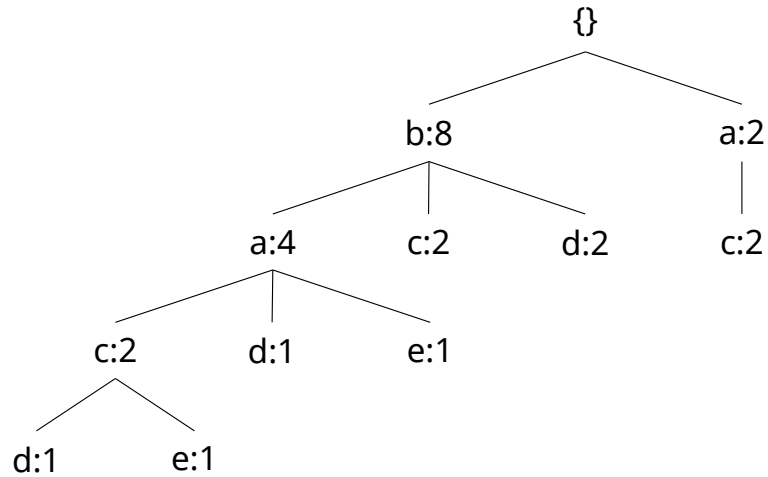
$$\{a, b, c\}, \{a, b, d\}, \{a, b, e\}$$

and any subset of these item sets.

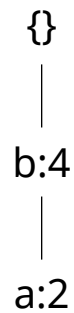
b) Our initial F-list contains a to e . If we arrange by frequency the list becomes $\{b, a, c, d, e\}$. The ordered frequent items in the database are shown in the following table.

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

The FP-tree is shown below.



c) The conditional pattern base for d is $\{bac : 1, ba : 1, b : 2\}$. We can see that only b and a are frequent here, so the conditional FP-tree is the one shown below.



d) Based on the conditional FP-tree we have the frequent patterns $\{bad : 2, ad : 2, bd : 4, d : 4\}$.

Problem 2

a)

b)

c)

Problem 3

a)

b)

Problem 4

a) This contains six elements. The length of s is eight. It contains $2^6 - 1 = 63$ non-empty subsequences.

b)