

~~Processing~~ Association rules

You are given the following list of transactions.

Apply the FP-growth algorithm using minsup = 2 (an itemset is frequent if it appears in at least two transactions).

Write the following:

1. B-CPB (B - Conditional Pattern Base)
2. B-CHT (B - Conditional Header Table)
3. BD-CPB (BD - Conditional Pattern Base)

Write all and only the correct itemsets, along with their support counts.

Use the following syntax:

```
B-CPB = { element1: support1, element2: support2, ... }
B-CHT = { element1: support1, element2: support2, ... }
BD-CPB = { element1: support1, element2: support2, ... }
```

For example:

$$B-CPB = \{ CDA: 1, CA: 1, A: 1, DA: 1, D \}$$

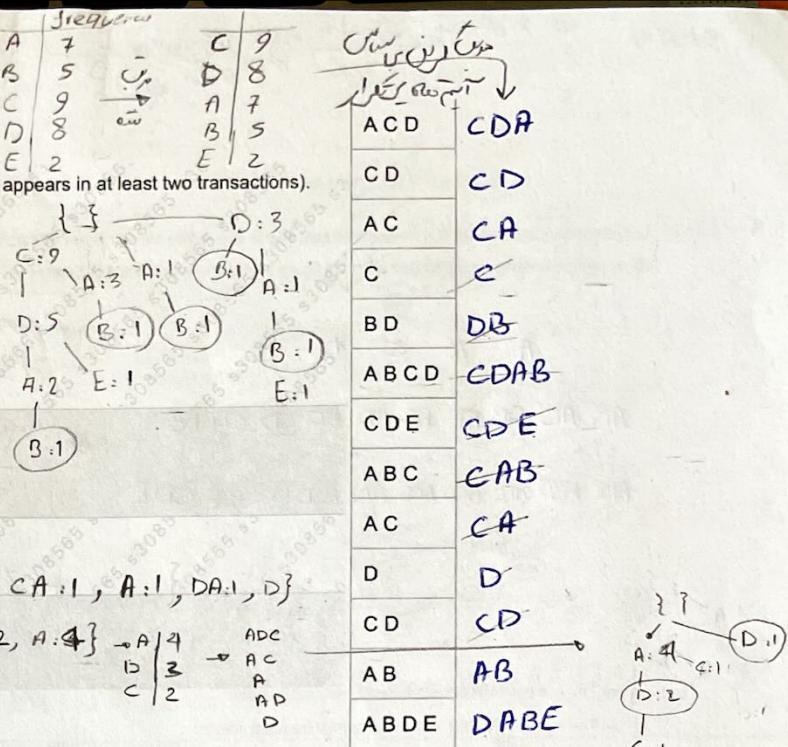
$$B-CHT = \{ a: 1, abc: 2 \}$$

$$B-CHT = \{ a: 3, b: 2, c: 1 \}$$

$$BD-CPB = \{ a: 1, b: 2, c: 3 \}$$

$$B-CHT = \{ C: 2, D: 2, A: 4 \}$$

$$BD-CPB = \{ A: 2 \}$$



An itemset is **closed** if none of its immediate supersets has the same support as the itemset.

Given the transactional dataset shown in the figure below, apply the **Apriori algorithm** to extract all frequent itemsets.

The value of minsup is 2.

An itemset is considered to be frequent if its support count is equal to or higher than the minsup.

Question 1) List all frequent itemsets having length 2, along with their support count.

Question 2) List all itemsets of length 3 that have been generated by Apriori after the join and prune steps, before counting their support in the database.

Question 3) List all frequent itemsets that are **not closed**, along with their support count.

Use the following notation:

- A1) { ... list of itemsets w/ support count ... }
- A2) { ... list of itemsets ... }
- A3) { ... list of itemsets w/ support count ... }

For example

- A1) { ab: 2, ac: 3, ad: 2 }
- A2) { abc, abd, abe }
- A3) { ab: 2, ad: 2, bce: 3 }

	Transactions
0	BE
1	BC
2	ABDE
3	BCDE
4	BC
5	ABE
6	BE
7	ACE
8	BDE
9	AB

You are given the following list of itemsets:

ADE	A: 7	AB	S	ABC	2
AB	B: 7	AC	2	ABD	2
BCD	C: 4	AD	3	ABE	3
BDE	D: 6	AE	5	AED	1
ABC	E: 6	BC	3	ACE	
AE		BD	4	AEE	
CD		BE	4	ADE	3
ABE		CE	1	BCD	2
ABCDE		CD	3	BCE	
ABDE		CE	1	BDE	3
		DE	4	CDE	

Apply the **Apriori algorithm** to extract all frequent itemsets, considering minsup = 2 (an itemset is frequent if it appears in at least two transactions). Only consider the extraction of itemsets of length up to 3 (included).

1. Which are the itemsets that can be pruned during the prune step (ie after the join step)?
2. Which are the itemsets that are pruned after calculating their support count?

Write the answer in the box below, using the following syntax:

1. (itemset1, itemset2, ..., itemsetN)
2. (itemset1, itemset2, ..., itemsetM)

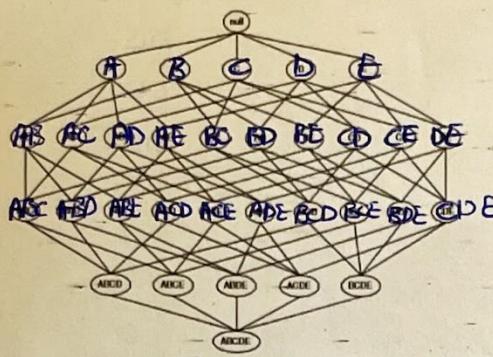
For example:

1. (A, B, AB)
2. (C, D, CD)

$$sup(x) = \frac{\text{ عدد المرة التي تظهر فيها } x}{\text{ عدد المرا }\times} \Rightarrow \text{Freshmax} \rightarrow x \text{ is frequent.}$$

~~Ques 10~~ A maximal frequent itemset is a frequent itemset whose immediate supersets are all infrequent.

Considering the itemsets generated as in the figure below, it is known that the itemset {ACE} is a maximal frequent itemset. Which of the following statements is correct?

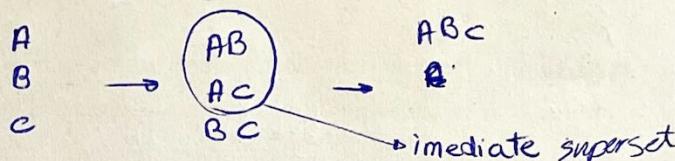


- (a) The itemset {ABCD} cannot be a frequent itemset
 - (b) The itemset {ACDE} is also a maximal frequent itemset
 - (c) None of the other statements are correct
 - (d) The itemset {ACD} cannot be a frequent itemset
 - (e) The itemset {BC} cannot be a frequent itemset
 - (f) The itemset {ACDE} is guaranteed to be a frequent itemset
 - (g) The itemset {ABCDE} cannot be a frequent itemset

- * an item is frequent maximal if none of its immediate superset is frequent.

~~This~~ ~~subset~~ is maxin frequent : ACE ^v maxin frequent set

* closed item → an Item is closed if none of its immediate subsets has the same support as the item.



Given the transactional dataset below, apply the Apriori algorithm to extract all frequent itemsets.

ABC 1
ABD
ABE o
ACD
ACE 1
ADE 2
BCE
BCE 2
BDE
CDF

For A
أيضاً \overline{AC} , AB أيضاً
وهو المثلث باز $\triangle ABC$ مغلق
غير مُنكمش \rightarrow باختصار
مثلاً $\triangle ABC$ مُنكمش
باختصار $\triangle ABC$ غير مُنكمش
باختصار $\triangle ABC$ مغلق

Case AB not closed (AB=30°) ABC=3 not closed ABC closed

• Eng not closed c AB open

The value of minsup is 2.

An itemset is considered to be frequent if its support count is equal to or higher than the minsup.

Question 1) List all frequent itemsets having length 2, along with their support count.

Question 2) List all itemsets of length 3 that have been generated by Apriori after the join and prune steps, before counting their support in the database.

Question 3) List all frequent itemsets that are **not closed**, along with their support counts.

Use the following notation:

- A1) { ... list of itemsets w/ support count ... }
- A2) { ... list of itemsets ... }
- A3) { ... list of itemsets w/ support count ... }

For example

- A1) { ab: 2, ac: 3, ad: 2 }
- A2) { abc, abd, abe }
- A3) { ab: 2, ad: 2, bce: 3 }

$r: A \rightarrow B$

$$\text{lift or correlation} = \frac{\text{conf}(r)}{\text{sup}(B)}$$

$a \rightarrow b$

$$\text{conf}(r) = \frac{\text{sup}(ab)}{\text{sup}(a)}, \text{ sup}(a) = \frac{\# a}{\# T} \rightarrow \text{support}$$

Given an association rule $r: A \rightarrow B$, lift(r) is defined as $\frac{\text{conf}(r)}{\text{sup}(B)}$, where conf(r) is the confidence of r, $\frac{\text{sup}(A,B)}{\text{sup}(A)}$ and sup(X) is the support of X.

Which of the following statements is correct?

(a) $\text{lift}(r) \in [-1, 1]$

$\text{conf}, \text{sup} \in [0, 1]$

$\text{lift} \in [0, +\infty]$

(b) if $\text{lift}(A \rightarrow B) > 1$, then $\text{lift}(A \rightarrow \text{not } B) < 1$

(c) None of the other statements are correct

(d) $\text{lift}(A \rightarrow B) \neq \text{lift}(B \rightarrow A)$

$$\frac{\text{sup}(a,b)}{\text{sup}(a)} = \frac{\text{sup}(a,b)}{\text{sup}(a)} \Rightarrow \text{lift}(A \rightarrow B) = \text{lift}(B \rightarrow A)$$

(e) $\text{lift}(r) > 0$ implies positive correlation

positive $\Leftrightarrow \text{conf}(r) > \text{sup}(B)$ i.e. $\text{lift}(r) > 1$

(f) $\text{lift}(r) < 0$ implies independence

independent $\Leftrightarrow \text{lift}(r) = 0$

The support of an itemset A (sup(A)) is the fraction of transactions that contains A and is represented by P(A).

The confidence of an association rule $A \rightarrow B$ is the conditional probability P(B|A).

The following rule $ab \rightarrow c$ with confidence($ab \rightarrow c$) = 1.0 is given. $\rightarrow \text{conf}(ab \rightarrow c) = \frac{\text{sup}(abc)}{\text{sup}(ab)} = 1$

Which statement is true?

(a) If a transaction contains ab, it will optionally contain c.

must

(b) If $\text{sup}(ab) = 1.0$, then c must be present in all transactions.

(c) None of the other answers is correct.

(d) $\text{sup}(c)$ can be 0.

(e) If $\text{sup}(c) = 1.0$, then $\text{sup}(ab)$ cannot take any value.

(f) $P(abc)$ must be 1.0

$\text{sup}(abc) = \text{sup}(ab)$ The support of an itemset A (sup(A)) is the fraction of transactions that contains A and is represented by P(A). $P(A) = \text{sup}(A)$

The confidence of an association rule $A \rightarrow B$ is the conditional probability $P(B|A) = \text{conf}(A \rightarrow B) = \frac{\text{sup}(AB)}{\text{sup}(A)}$

The following rule $ab \rightarrow c$ with confidence($ab \rightarrow c$) = 1.0 is given. $\text{conf}(ab \rightarrow c) = \frac{\text{sup}(abc)}{\text{sup}(ab)} = 1$

Which statement is true?

(a) None of the other answers is correct.

(b) If $\text{sup}(c) = 1.0$, then $\text{sup}(ab)$ cannot take any value.

$\text{sup}(c) = 1 \Rightarrow \text{sup}(a,b) = 1$ i.e. $\text{conf}(ab \rightarrow c) = 1$

(c) $P(abc)$ must be 1.0.

(d) If $\text{sup}(ab) = 1.0$, then c must be present in all transactions.

(e) $\text{sup}(c)$ can be 0.

(f) If a transaction contains ab, it will optionally contain c.

must

Given an association rule $r: A \rightarrow B$, lift(r) is defined as $\frac{\text{conf}(r)}{\text{sup}(B)}$, where conf(r) is the confidence of r , $\frac{\text{sup}(A, B)}{\text{sup}(A)}$ and sup(X) is the support of X .

Which of the following statements is correct?

- (a) $\text{lift}(r) > 0$, implies positive correlation $\text{lift}(r) > 1 \Rightarrow \text{positive corr}$

- (b) None of the other statements is correct

- (c) $\text{lift}(A \rightarrow B) \neq \text{lift}(B \rightarrow A)$

$$\frac{\frac{\text{sup}(ab)}{\text{sup}(a)}}{\frac{\text{sup}(b)}{\text{sup}(a)}} = \frac{\text{sup}(ab)}{\text{sup}(a) \cdot \text{sup}(b)}$$

- (d) $\text{lift}(r) \in [-1, 1] \rightarrow \text{lift} \in [0, +\infty)$

- (e) $\text{lift}(r) < 0$ implies independence $\text{if } \text{lift} = 0 \Rightarrow \text{independence}$

- (f) if $\text{lift}(A \rightarrow B) > 1$, then $\text{lift}(A \rightarrow \text{not } B) < 1$

$$\text{lift}(A \rightarrow \bar{B}) < 1 \Leftrightarrow \text{not } B \text{ is independent from } A \text{ (میان رسانی بین } A \text{ و } \bar{B} \text{ نداشته باشد)}$$

$A \rightarrow B$

The support of a set of items A ($\text{sup}(A)$) is the fraction of transactions that contains A and is represented by $P(A)$.

The confidence of an association rule $A \rightarrow B$ is the conditional probability $P(B|A)$.

The following rule $ab \rightarrow c$ with confidence($ab \rightarrow c$) = 1.0 is given. $\Rightarrow \frac{\text{sup}(abc)}{\text{sup}(ab)} = 1 \Rightarrow \text{sup}(abc) = \text{sup}(ab)$

Which statement is true?

- (a) If $\text{sup}(c) = 1.0$, then $\text{sup}(ab)$ cannot take any value.

- (b) If $\text{sup}(ab) = 1.0$, then c must be present in all transactions.

- (c) If a transaction contains ab , it will optionally contain c .

- (d) None of the other answers is correct.

- (e) $P(abc)$ must be 1.0. $\rightarrow \text{sup}(ab) = 1 \rightarrow \text{sup}(abc) = 1$

- (f) $\text{sup}(c)$ can be 0. $\rightarrow \frac{\text{sup}(abc)}{\text{sup}(ab)} = 1 \Rightarrow \text{sup}(c) \neq 0 \rightarrow \text{sup}(c) \neq 0$

An itemset I is closed if none of its immediate supersets has the same support as I .

An itemset I is represented by a collection of literals (e.g. abc) and a number representing its support count (e.g. abc: 12).

After the analysis of a transactional dataset, the list of frequent itemsets found (with support counts) is the following:

length=1 a:150 b:160 c:180 d:150 e:150

length=2 ab:140 ac:120 ad:130 ae:140

length=3 abc:120 acd:110 abe:130

length=4 abce:120

Which of the following statements is true?

not closed $\rightarrow \text{sup}(I) = \text{sup}(\text{immediate})$

! is it closed? Or, To immediate
how come? \leftarrow not closed

- (a) The only non-closed itemsets are ac, ad

- (b) The only non-closed itemsets are ac, abc

- (c) The only non-closed itemsets are ac, ad, abc, abce

- (d) The only non-closed itemsets are ac, ad

- (e) The only closed itemsets are ac, abc

- (f) None of the other answers is correct

The confidence of a rule is defined as:

$$\text{confidence}(A \rightarrow B) = \frac{\text{sup}(A, B)}{\text{sup}(A)}$$

where $\text{sup}(x)$ indicates the fraction of transactions where x occurs (i.e. the support).

Let R: $a, b \rightarrow c, d$ be an association rule. r: $ab \rightarrow cd$

In the transactional database you find the following records:

a b c d
a e f
a c d
a b
b c d f
b c d e
a e
b f

$$\text{conf}(ab \rightarrow cd) = \frac{\text{sup}(abcd)}{\text{sup}(ab)} = \frac{1}{2}$$

- (a) $\text{confidence}(R) = 8/6$
- (b) $\text{confidence}(R) = 24/25$
- (c) $\text{confidence}(R) = 2/3$
- (d) $\text{confidence}(R) = 1/3$
- (e) None of the answers is correct
- (f) $\text{confidence}(R) = 1/2$

Which of the following statements is true?

An itemset is defined to be frequent maximal if none of its immediate supersets (i.e. including one more item) is frequent.

After the analysis of a transactional dataset, the complete set of maximal frequent itemsets with minsup=900 are:

abc: 930 $\rightarrow c \in [930, +\infty)$

def: 980 $\rightarrow de \in [980, +\infty)$

Which of the following statements is true?

- (a) $\text{sup}(ab)$ can be in range [900, 929], $\text{sup}(abcd)$ cannot be in range [0, 930]
- (b) $\text{sup}(ab)$ can be a value in range [900, 929], $\text{sup}(cd)$ cannot be in range [900, 930]
- (c) $\text{sup}(ab)$ can be a value in range [930, 1000], $\text{sup}(abcd)$ can be a value in range [900, 930] $\rightarrow X$
- (d) None of the other answers is correct
- (e) $\text{sup}(ab)$ can be a value in range [930, 1000], $\text{sup}(cd)$ cannot be in range [900, 930]
- (f) $\text{sup}(ab)$ can be a value in range [900, 929], $\text{sup}(abcd)$ can be a value in range [900, 930]

$$abcd < 900 \rightarrow cd < 900 ?$$

- cd frequent

$$cd \in [930, +\infty)$$

a: 6	AB: 3	BC: 2	ABC: 1	BCD: 1
b: 9	AC: 4	BD: 0	ABD: 1	BCE: 1
c: 6	AD: 2	BE: 1	ABE: 1	BDE: 1
D: 5	AE: 2	CD: 3	ACD: 2	CDE: 2
E: 5	CE: 2	DE: 1	ACE: 2	

Given the transactional dataset shown in the figure below, apply the Apriori algorithm to extract all frequent itemsets.

The value of minsup is 2.

An itemset is considered to be frequent if its support count is equal to or higher than the minsup.

Question 1) List all frequent itemsets having length 2 along with their support count. {AB:3, AC:4, AD:2, AE:2, BC:2, CD:3, CE:3, DE:1}

Question 2) List all itemsets of length 3 that have been generated by Apriori after the join and prune steps, before counting their support on the database. {ABC:1, ACD:2, ACE:2, ADE:1, CDE:2}

Question 3) List all frequent itemsets that are NOT closed. along with their support count. {AD:2, AE:2}

You are given the following list of transactions.

dc	5	d	9
cd	7	b	7
dbae	6	c	6
abde	9	a	5
bcde	3	e	3
dbc	30		
bca			
abc			
d			
abcd			
dc			
abde			
bd			
db			

Write the following:

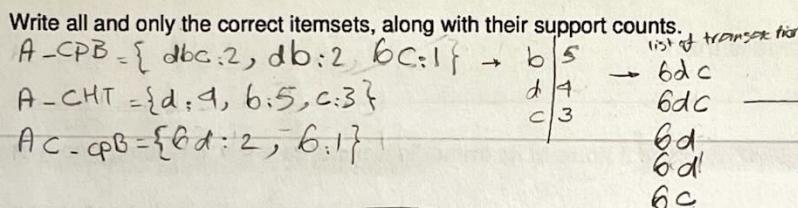
1. A-CPB (A - Conditional Pattern Base)
2. A-CHT (A - Conditional Header Table)
3. AC-CPB (AC - Conditional Pattern Base)

Write all and only the correct itemsets, along with their support counts.

$$A\text{-CPB} = \{dbc:2, db:2, bc:1\}$$

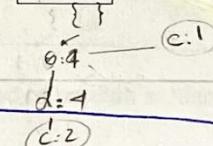
$$A\text{-CHT} = \{d:4, b:5, c:3\}$$

$$AC\text{-CPB} = \{bd:2, b:1\}$$



Transactions

ab
cde
bc
ab
acd
abce
de
acde
ac
de



Apply the FP-growth algorithm using minsup = 2 (an itemset is frequent if it appears in at least two transactions).

- (a) sup(B) < 1/2
- (b) Not enough information is provided to address the question
- (c) None of the other answers is correct
- (d) sup(B) > 1/2
- (e) sup(B) > 1/10
- (f) sup(B) > 1/5
- (g) sup(B) < 1/5
- (h) sup(B) < 1/10

For a rule $A \rightarrow B$, the following can be computed:

$$\text{confidence}(A \rightarrow B) = \frac{\text{sup}(A, B)}{\text{sup}(A)}$$

$$\text{lift}(A \rightarrow B) = \text{correlation}(A \rightarrow B) = \frac{\text{confidence}(A \rightarrow B)}{\text{sup}(B)}$$

Where sup(X) is the support of an itemset X, i.e. the fraction of transactions that contain X.

A frequent itemset is maximal if none of its supersets is frequent.

You are given a transactional dataset. You extract all frequent itemsets using a minimum support of 0.1 (an itemset is considered frequent if it appears in at least 10% of the transactions).

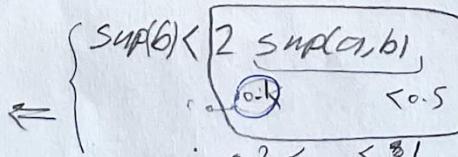
You find that the itemset A = {a1, a2, ...} has support $\text{sup}(A) = 0.5$, and it is maximal.

Which of the following constraints on the support of B = {b1, b2, ...} guarantees that the rule $A \rightarrow B$ has positive correlation?

$$\frac{\text{sup}(a, b)}{\text{sup}(a) \cdot \text{sup}(b)} > 1 \rightarrow 2 \text{sup}(a, b) > \text{sup}(b)$$

If applicable (i.e. a solution can be found and is among the provided answers), choose the most permissive constraint, which is the constraint that produces the largest possible range of permissible values. For example, if both $\text{sup}(B) > x$ and $\text{sup}(B) > y$ satisfy the request, you must choose $\text{sup}(B) > \min(x, y)$. Likewise if the options are in the form $\text{sup}(B) < x$ and $\text{sup}(B) < y$, choose $\text{sup}(B) < \max(x, y)$.

$$\text{sup}(b) < 0.2$$



$$\text{sup}(B) = \frac{\#B}{\#T}$$

0.1

$$\text{sup}(B) < \text{sup}(a)$$

Closed, frequent maximal

most frequent is maximal

An itemset is maximal if it is frequent and none of its supersets is frequent

After extracting the frequent itemsets from a transactional dataset, you find that -- among others -- the itemsets ABC and BD are maximal.

Which of the following statements is correct?

- (a) AC is frequent ✓
- (b) ABCD is frequent
- (c) ABD is maximal
- (d) AC is maximal
- (e) ABCD is frequent only if BCD is frequent
- (f) None of the other statements is correct
- (g) ABD is frequent
- (h) ABCD is maximal

D E Given the transactional dataset shown in the figure below, apply the Apriori algorithm to extract all
AC D E frequent itemsets.

A B C E

A B C

C E

A C D

C D

A C D

A C D E

B D

The value of minsup is 2 (an itemset is frequent if it appears in at least 2 transactions).

An itemset is considered to be frequent if its support count is equal to or higher than the minsup.

1. List all frequent itemsets having length 2, along with their support count.
2. List all candidate itemsets of length 3 that have been generated by Apriori after the join and prune steps, before counting their support in the database.

3. List all frequent itemsets that are not closed, along with their support count. ?

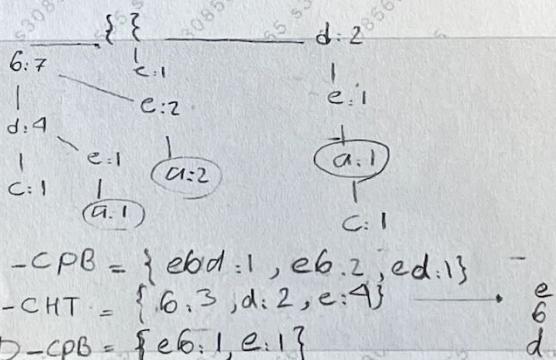
? $\times 2. \{ABC, ACD, ACE, ADE, CDE\}$

3. $\{ABC:2, ACD:4,$

3 points (no penalty for a wrong answer)

You are given the following list of transactions.

Item	Frequency
bcd	6dc
acde	deac
bd	bd
b	b
abe	bea
e	e
abe	6ea
bd	bd
d	d
abde	bd ea



$$A - CPB = \{ebd:1, e6:2, ed:1\}$$

$$A - CHT = \{b:3, d:2, e:4\}$$

$$AD - CPB = \{eb:1, e:1\}$$

A: 6	AB: 2	ABC: 2
B: 3	AC: 6	ACD: 4
C: 8	AD: 4	ACE: 3
D: 7	AE: 3	ADE: 2
E: 5	BC: 2	BCD: 1
	BD: 1	BE: 1
	CD: 4	CE: 4
	DE: 3	ODE: 2

Apply the FP-growth algorithm using minsup = 2 (an itemset is frequent if it appears in at least two transactions).

Write the following:

1. A-CPB (A - Conditional Pattern Base)
2. A-CHT (A - Conditional Header Table)
3. AD-CPB (AD - Conditional Pattern Base)

Write all and only the correct itemsets, along with their support counts.

Use the following syntax:

A-CPB = { element1: support1, element2: support2, ... }

A-CHT = { element1: support1, element2: support2, ... }

AD-CPB = { element1: support1, element2: support2, ... }

For example:

A-CPB = { a: 1, abc: 2 }

A-CHT = { a: 3, b: 2, c: 1 }

AD-CPB = { a: 1, b: 2, c: 3 }

