

Frequent itemsets

$\text{minsup} = 3$

D	A	B	C	D	E
1	1	1	0	1	1
2	0	1	1	0	1
3	1	1	0	1	1
4	1	1	1	0	1
5	1	1	1	1	1
6	0	1	1	1	0

t	$\mathbf{i}(t)$
1	ABDE
2	BCE
3	ABDE
4	ABCE
5	ABCDE
6	BCD

$$\mathcal{F}^{(1)} = \{A, B, C, D, E\}$$

$$\text{sup}(AB) = 4 \geq 3 \checkmark$$

$$\text{sup}(AC) = 2 \neq 3 \times$$

$$\mathcal{F}^{(2)} = \{AB, AD, \dots\}$$

$$\text{sup}(ABCDE) = 1 \leq 3 \times$$

Do we need to check this itemset

Association rules

Given disjoint itemsets X, Y

association rule

$$X \rightarrow Y \quad \{\text{milk, eggs}\} \rightarrow \{\text{cereal}\}$$

* notation $X \cup Y$ write as XY

support of a rule

$$\text{sup}(X \rightarrow Y) = |t(XY)| = \text{sup}(XY)$$

rel support
of rule

$$\text{rsup}(X \rightarrow Y) = \frac{|t(XY)|}{|D|} = \frac{\text{sup}(XY)}{\text{sup}(X \cup Y)} = \text{rsup}(XY)$$

(est of joint prob of $X \cup Y$)

confidence of rule

$$\text{conf}(X \rightarrow Y) = P(Y|X) = \frac{P(X \cap Y)}{P(X)} = \frac{\text{sup}(XY)}{\text{sup}(X)}$$

note: prob of an itemset
w/ items in X and items in Y

Given

rule $X \rightarrow Y$

$\text{minsup} \in \mathbb{Z}^+$

(user defined)

$\text{minconf} \in [0, 1]$

(user defined)

-rule is frequent if $\text{sup}(X \rightarrow Y) \geq \text{minsup}$

-rule is Strong if $\text{conf}(X \rightarrow Y) \geq \text{minConf}$

XY is freq

Itemset and rule mining

prob^I itemset mining

Given a binary database D
minimum support threshold minsup (\geq^+)
Find all frequent itemsets in D

prob II

Given frequent itemset F
min confidence minconf ($[0,1]$)
Find all frequent and strong rules

Algo For Frequent itemset (Brute force algo)

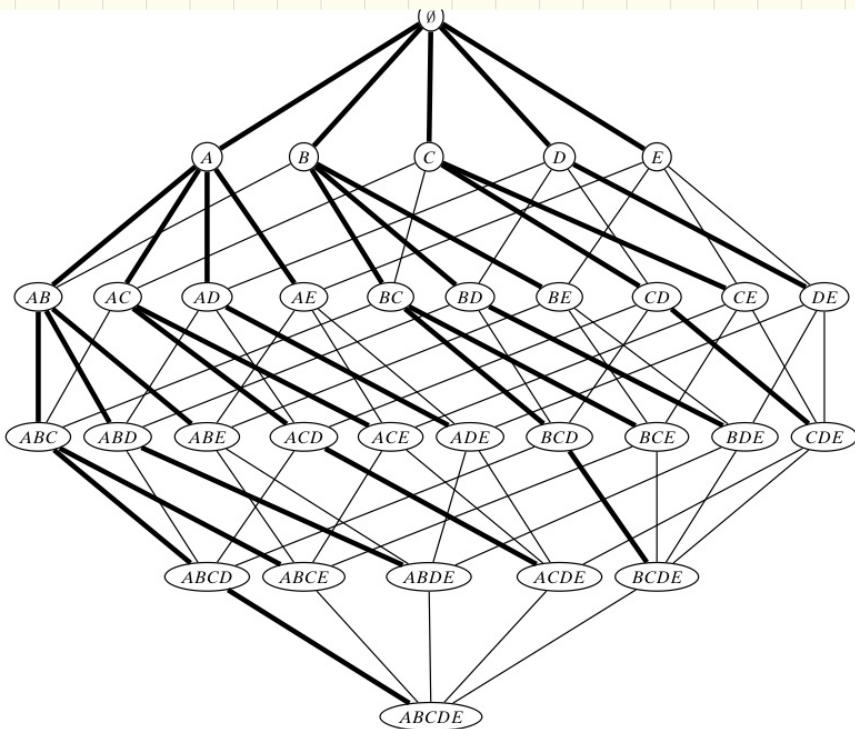
COMPUTESUPPORT (X, D):

```
7  $sup(X) \leftarrow 0$ 
8 foreach  $\langle t, i(t) \rangle \in D$  do
9   | if  $X \subseteq i(t)$  then
10    |   |  $sup(X) \leftarrow sup(X) + 1$ 
11 return  $sup(X)$ 
```

Algorithm 8.1: Algorithm BRUTEFORCE

BRUTEFORCE ($D, I, minsup$):

```
1  $\mathcal{F} \leftarrow \emptyset$  // set of frequent itemsets
2 foreach  $X \subseteq I$  do
3   |  $sup(X) \leftarrow \text{COMPUTESUPPORT}(X, D)$ 
4   | if  $sup(X) \geq minsup$  then
5     |   |  $\mathcal{F} \leftarrow \mathcal{F} \cup \{(X, sup(X))\}$ 
6 return  $\mathcal{F}$ 
```



COMPUTESUPPORT (X, \mathbf{D}):

- 7 $sup(X) \leftarrow 0$
- 8 **foreach** $\langle t, i(t) \rangle \in \mathbf{D}$ **do**
- 9 **if** $X \subseteq i(t)$ **then**
- 10 $sup(X) \leftarrow sup(X) + 1$
- 11 **return** $sup(X)$

$$\text{Recall } \sup(X, \mathbf{D}) = \left| \left\{ t \mid (t, i(t)) \in \mathbf{D} \text{ and } X \subseteq i(t) \right\} \right| = |t(x)|$$

sketch

for $t \in T$
 for $l \in i(t)$ $\boxed{\quad}$ $O(|D|)$
 for $x \in X$ $O(|I^x|)$
 does $l = x$ $O(1)$

support takes $O(|I^x| |D|)$ work

Algorithm 8.1: Algorithm BRUTEFORCE

BRUTEFORCE ($\mathbf{D}, I, minsup$):

- 1 $\mathcal{F} \leftarrow \emptyset$ // set of frequent itemsets
- 2 **foreach** $I \subseteq I$ **do**
- 3 $sup(I) \leftarrow \text{COMPUTESUPPORT}(I, \mathbf{D})$
- 4 **if** $sup(I) \geq minsup$ **then**
- 5 $\mathcal{F} \leftarrow \mathcal{F} \cup \{(I, sup(I))\}$
- 6 **return** \mathcal{F}

How many times do we need to compute the support?
 - # of itemsets = $2^{|I|}$
 \Rightarrow total work w/ brute force algo
 is $O(2^{|I|} |I| |D|)$

I/O complexity

$2^{|I|}$ DB scans per support call
 calls to support $\Rightarrow 2^{|I|}$ passes over the DB
 $\Rightarrow O(2^{|I|}) = \Theta(2^{|I|})$ passes over the DB