

CS620

HW5

@author: Joseph Cannella

Homework #5: Pattern Mining

1.) Given: Table (T)

a.) Compute Support + support count for itemsets
 $\{Asprin\}$, $\{Tylenol, Cepacol\}$, $\{Asprin, Ibuprofen, Panadol\}$

Itemset	Support Count	Support	$N=12$
$\{Asprin\}$	6	$6/12 = 0.5$	
$\{Tylenol, Cepacol\}$	4	$4/12 = 0.33\bar{3}$	
$\{Asprin, Ibuprofen, Panadol\}$	1	$1/12 = 0.08\bar{3}$	

b.) Find confidence for following association rules:

• Confidence (C) - how often items in Y appear in transactions that contain X

"How often rule is true"

Rule	$N(X)$	$N(Y)$	Confidence (C)
$\{Asprin, Vitamin C\} \Rightarrow \{Swlated\}$	2	1	$1/2 = 0.5$
$\{Asprin\} \rightarrow \{Vitamin C\}$	6	2	$2/6 = 0.33\bar{3}$
$\{Vitamin C\} \rightarrow \{Asprin\}$	3	2	$2/3 = 0.66\bar{6}$

Last two results are different because

Asprin is more common than Vitamin C,
but $N(Asprin + Vitamin C)$ are the same
in both confidence calculations.

Homework #5: Pattern Mining

items
1 a, c, d, e
2 a, d, e, f
b, c, d, e, f
b, d, e, f
b, e, f
c, d, e
c, e, f
d, e, f

a.) Rule	$N(X)$	$N(Y)$	Confidence (c)
$\{a\} \rightarrow \{f\}$	2	1	$1/2 = 0.5$
$\{a, e\} \rightarrow \{f\}$	2	1	$1/2 = 0.5$

b.) $\text{minsup} =$

Homework #5: Pattern Mining

2-b) Apriori Algorithm, minsup = 3

Level	Itemset	Support	Is Valid
1	A	2	X
	B	3	✓
	C	4	✓
	D	6	✓
	E	8	✓
	F	6	✓
2	BC	1	X
	BD	2	X
	BE	3	✓
	BF	3	✓
	CD	3	✓
	CE	4	✓
	CF	2	X
	DE	6	✓
	DF	4	✓
3	BFF	3	✓
	CDE	3	✓
	CDF	1	X
	DEF	4	✓
	EF	6	✓

Homework #5: Pattern Mining

2-c) Association Rules, Confidence = 80%

Frequent Item sets:

L1: $\{B\}, \{C\}, \{D\}, \{E\}, \{F\}$

L2: $\{BE\}, \{BF\}, \{CD\}, \{CE\}, \{DE\}, \{DF\}, \{EF\}$

L3: $\{BEF\}, \{CDE\}, \{DEF\}$

L2: Rules	Support (XY)	Support (X)	Confidence
$\{B\} \rightarrow \{E\}$	3	3	100%
$\{E\} \rightarrow \{B\}$	3	8	37.5%
$\{B\} \rightarrow \{F\}$	3	3	100%
$\{F\} \rightarrow \{B\}$	3	6	50%
$\{C\} \rightarrow \{D\}$	3	4	75%
$\{D\} \rightarrow \{C\}$	3	6	50%
$\{C\} \rightarrow \{E\}$	4	4	100%
$\{E\} \rightarrow \{C\}$	4	8	50%
$\{D\} \rightarrow \{E\}$	6	6	100%
$\{E\} \rightarrow \{D\}$	6	8	75%
$\{D\} \rightarrow \{F\}$	4	6	66.7%
$\{F\} \rightarrow \{D\}$	4	6	66.7%
$\{E\} \rightarrow \{F\}$	6	8	75%
$\{F\} \rightarrow \{E\}$	6	6	100%

Cont'd →

Homework #2: Pattern Mining

2-C) L3: {BEF}, {CDE}, {DEF}

L3 Rules	Support(XY)	Support(X)	Confidence (C)
$\{B\} \rightarrow \{EF\}$	3	3	100%
$\{EF\} \rightarrow \{B\}$	3	6	50%
$\{BE\} \rightarrow \{F\}$	3	3	100%
$\{F\} \rightarrow \{BE\}$	3	6	50%
$\{BF\} \rightarrow \{E\}$	3	3	100%
$\{E\} \rightarrow \{BF\}$	3	8	37.5%
$\{C\} \rightarrow \{DE\}$	3	4	75%
$\{DE\} \rightarrow \{C\}$	3	6	50%
$\{CD\} \rightarrow \{E\}$	3	3	100%
$\{E\} \rightarrow \{CD\}$	3	3	100%
$\{CE\} \rightarrow \{D\}$	3	3	100%
$\{D\} \rightarrow \{CE\}$	3	6	50%
$\{D\} \rightarrow \{EF\}$	4	6	66.7%
$\{EF\} \rightarrow \{D\}$	4	6	66.7%
$\{E\} \rightarrow \{DF\}$	4	8	50%
$\{DF\} \rightarrow \{E\}$	4	4	100%
$\{F\} \rightarrow \{DE\}$	4	6	66.7%
$\{DE\} \rightarrow \{F\}$	4	6	66.7%