Association rules

Association rules are rules of the form **X → Y**, where X and Y are conjunctions of items. Task: Find **all** association rules that satisfy minimum support and minimum confidence constraints.

Support:	$Sup\ (X \rightarrow Y) = \#XY\ /\ \#D \cong p(XY)$
Confidence:	$Conf\ (X \rightarrow Y) = \#XY\ /\ \#X \cong p(XY)\ /\ p(X) = p(Y|X)$

Association rules algorithm *Apriori* (data, minSupport, minConfidence)

1. Generate frequent itemsets with a minimum support constraint.
2. Generate rules from frequent itemsets with a minimum confidence constraint.

Frequent itemsets mining

Items in an itemset should **always** be sorted alphabetically.
- Generate all 1-itemsets with the given minimum support.
- Use 1-itemsets to generate 2-itemsets with the given minimum support.
- From 2-itemsets generate 3-itemsets with the given minimum support as unions of 2-itemsets with the same item at the beginning.
- …
- From n-itemsets generate (n+1)-itemsets as unions of n-itemsets with the same (n-1) items at the beginning.
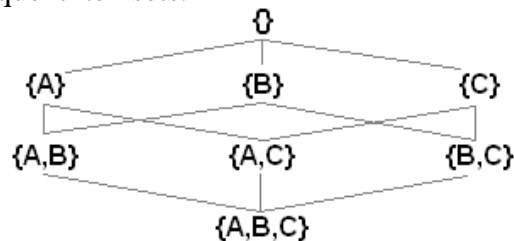
The result is a lattice of frequent itemsets.

Figure 1: A lattice of items A, B and C

Generating rules from frequent itemsets

From each frequent itemset construct all possible rules.
The support of the rule is the same as the support of the itemset it was constructed from.
Calculate the confidence of each rule. All the counts are in the lattice.

Exercise

Find all association rules with minimum support 2/6 and minimum confidence 70%.

A	B	C	D
1	1	1	1
	1	1	
	1		1
1		1	
1	1		1
1	1	1	

A = apple
B = banana
C = coca-cola
D = doughnut

Petra Kralj
Petra.Kralj@ijs.si
http://kt.ijs.si/PetraKralj/IPSKnowledgeDiscovery0708.html