

# Association Analysis

## Exercise 8



# Frequent Itemsets & Association Rules

- Frequent Itemset
  - Support count: Frequency of an itemset
  - Support: relative frequency of an itemset (wrt. all transactions)
- Association Rule  $X \rightarrow Y$ 
  - Support: Support of the itemset  $X \cup Y$
  - Confidence: relative frequency of  $X \cup Y$  wrt.  $X$ 
    - “If an itemset contains  $X$ , in  $x\%$  of the cases it also contains  $Y$ ”
  - Lift: confidence of rule  $X \rightarrow Y$  divided by support of consequent  $Y$ 
    - $>1$       $X$  and  $Y$  are positively correlated
    - $<1$       $X$  and  $Y$  are negatively correlated
    - $=1$       $X$  and  $Y$  are independent

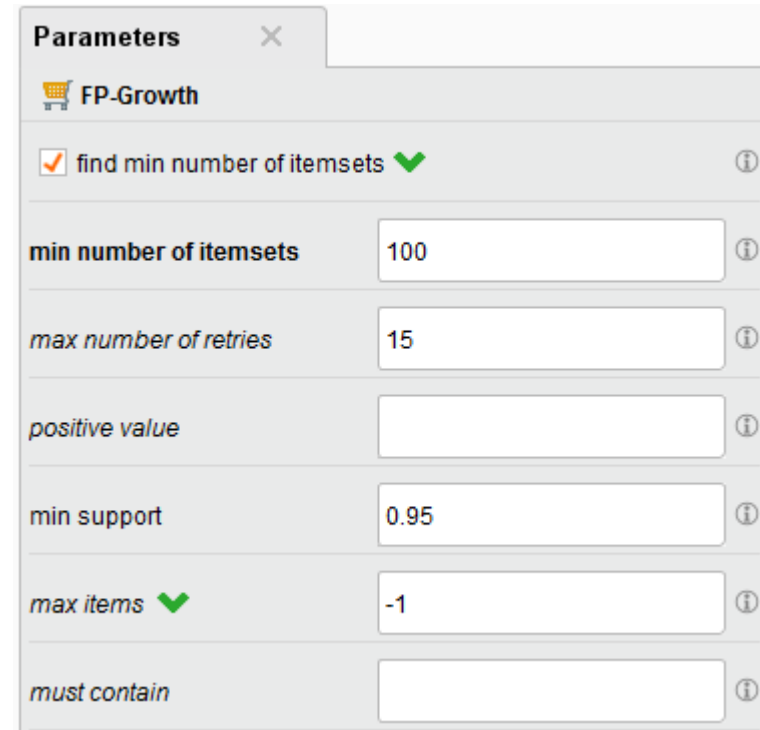
# The Apriori Principle

- If an itemset is frequent, then all of its subsets must also be frequent
  - Support of an itemset never exceeds the support of its subsets
  - “anti-monotone” property of support
- Example:
  - 50% of all shopping baskets contain beer
    - Then beer & nachos can’t be in more than 50% of all shopping baskets
  - So if any itemset is less frequent than we require, no need to check larger itemsets

$$\forall X, Y : (X \subseteq Y) \Rightarrow s(X) \geq s(Y)$$

# Operators: FP-Growth

- Input Port
  - Example Set (binominal attributes!)
- Output Ports
  - Example Set
  - Frequent Item Sets
- Parameters
  - If “find min number of itemsets”
    - Min number of itemsets
  - Else
    - Min support
  - Positive value
  - Max items
  - Must contain




The screenshot shows a 'Parameters' dialog box for the 'FP-Growth' operator. It features a shopping cart icon and the text 'FP-Growth'. A checked checkbox labeled 'find min number of itemsets' is followed by a green checkmark and an information icon. Below this are several input fields, each with an information icon to its right: 'min number of itemsets' with the value '100', 'max number of retries' with the value '15', 'positive value' (empty), 'min support' with the value '0.95', 'max items' with a green checkmark and the value '-1', and 'must contain' (empty).



# Operators: Create Association Rules

- Input Port
  - Frequent Item Sets
- Output Ports
  - Association Rules
  - Frequent Item Sets
- Parameters
  - Criterion
  - Min criterion value
  - Gain theta (used if criterion = gain)
  - Laplace k (used if criterion = laplace)

Parameters ×

 Create Association Rules

critierion	confidence	<span>⌵</span> ⓘ
min confidence	0.8	ⓘ
gain theta	2.0	ⓘ
laplace k	1.0	ⓘ

