香港城市大學
City University of Hong Kong

專業 創新 胸懷全球
Professional·Creative
For The World

# Frequent Pattern Analysis:
## Apriori Algorithm

CS5483 Data Warehousing and Data Mining

# Market basket analysis

| Transaction ID (TID) | Purchased Items |
|---|---|
| 1 | Tissue paper, bread, milk |
| 2 | Tissue paper, bread, milk |
| 3 | Tissue paper, bread, jam |
| 4 | Tissue paper, bread, jam |
| 5 | Tissue paper, milk, diaper |

- Supermarket may want to:
  - Increase the stocks of **frequently purchased items**.
  - Bundle together items that are *frequently purchased together*.
- How to learn from transactional data?

# Problem formulation

| Transaction ID (TID) | Purchased Items |
|---|---|
| 1 | Tissue paper, bread, milk |
| 2 | Tissue paper, bread, milk |
| 3 | Tissue paper, bread, jam |
| 4 | Tissue paper, bread, jam |
| 5 | Tissue paper, milk, diaper |

- $\mathcal{I} \coloneqq \{I_1, \dots, I_m\}$: Set of **items**, e.g.,
  $\mathcal{I} = \{\text{tissue paper,bread,milk,jam,diaper}\}$
- $A \subseteq \mathcal{I}$: **I**____________, written as $\mathrm{A} = \{I_{j_1}, I_{j_2}, \dots\}$ with $1 \leq j_1 < j_2 < \cdots \leq m$.
- **Transactional data**: $D \coloneqq \{T_i\}_{i=1}^{n}$ where $T_i \subseteq \mathcal{I}$, e.g.,
  $T_1 = \{\text{tissue paper,bread,milk}\}$

# Problem formulation

| Transaction ID (TID) | Purchased Items |
|---|---|
| 1 | Tissue paper, bread, milk |
| 2 | Tissue paper, bread, milk |
| 3 | Tissue paper, bread, jam |
| 4 | Tissue paper, bread, jam |
| 5 | Tissue paper, milk, diaper |

- **(Support) count** of itemset $A \subseteq \mathcal{J}$ in $D$:

$$\text{count}(A) := |\{T \in D | A \subseteq T\}|$$

- Objective: Given $\text{min\_sup} > 0$, obtain the set of **f**____________ **$k$-itemsets**

$$L_k := \{A \subseteq \mathcal{J} | |\text{A}|\text{=k}, \text{count}(A) \geq \text{min\_sup}\}, \qquad \text{for all } k \in \{1, \dots, m\}.$$

# Example

$$\begin{array}{llllll}
T_1 = & \{ & & I_3, & I_4, & I_5\} \\
T_2 = & \{ & & I_3, & I_4, & I_5\} \\
T_3 = & \{ & I_2, & & I_4, & I_5\} \\
T_4 = & \{ & I_2, & & I_4, & I_5\} \\
T_5 = & \{I_1, & & I_3, & & I_5\}
\end{array}$$

- $\mathcal{I} = \{I_1, I_2, I_3, I_4, I_5\}$ and $D = \{T_1, T_2, T_3, T_4, T_5\}$.
- $\text{min\_sup} = 2$.
  - $\{I_2, I_5\}$ is a frequent itemset. Any others? ______________________________
  - $\{I_2, I_3, I_5\}$ is not a frequent itemset. Any others? ________________________

# Compute frequent 1-itemsets

$$\begin{array}{llllllll} T_1 & = & \{ & & & 3, & 4, & 5\} \\ T_2 & = & \{ & & & 3, & 4, & 5\} \\ T_3 & = & \{ & & 2, & & 4, & 5\} \\ T_4 & = & \{ & & 2, & & 4, & 5\} \\ T_5 & = & \{ & 1, & & 3, & & 5\} \end{array}$$

with $I_j \coloneqq j$ for notational simplicity.

$C_1$

| | |
|---|---|
| {1} | |
| {2} | |
| {3} | |
| {4} | |
| {5} | |

$L_1$

| | |
|---|---|
| {2} | |
| {3} | |
| {4} | |
| {5} | |

- Generate a **c**_______________ **list** $C_1$ consisting of singleton sets.
- Look up the count of every $\mathrm{A} \in C_1$
- Add $A$ from $C_1$ to $L_1$ if count$(A) \geq$ min_sup.

# Compute frequent 2-itemsets

$T_1 = \{\quad 3,\ 4,\ 5\}$

$T_2 = \{\quad 3,\ 4,\ 5\}$

$T_3 = \{\ 2,\quad 4,\ 5\}$

$T_4 = \{\ 2,\quad 4,\ 5\}$

$T_5 = \{1,\quad 3,\quad 5\}$

$C_2$

| | |
|---|---|
| {2,3} | |
| {2,4} | |
| {2,5} | |
| {3,4} | |
| {3,5} | |
| {4,5} | |

$L_2$

| | |
|---|---|
| {2,4} | |
| {2,5} | |
| {3,4} | |
| {3,5} | |
| {4,5} | |

$L_1$

| | |
|---|---|
| {2} | |
| {3} | |
| {4} | |
| {5} | |

- Generate the candidate list $C_2$ of 2-itemsets with each item taken from $L_1$.
- Why not take from $\mathcal{I}$ instead of $L_1$? (i.e., why not consider item 1?)

# Apriori property

- count($A$) is **anti**_________________________:

  $$\text{count}(A) \leq \text{count}(B), \qquad \text{for any non−empty } B \subseteq A \subseteq \mathcal{I}$$

  - All non-empty subsets of a frequent itemset must be frequent/infrequent.
  - All supersets of a infrequent itemset must be frequent/infrequent.
- Other such functions?
  - Minimum value of items in $A$. Y/N
  - Maximum value of items in $A$. Y/N
  - Total value of items in $A$. Y/N
  - $A \mapsto \sum_{T \in D : A \subseteq T} x_T$ when $x_T$'s are non-negative. Y/N

# Compute frequent 3-itemsets

$T_1 = \{\quad 3,\ 4,\ 5\}$

$T_2 = \{\quad 3,\ 4,\ 5\}$

$T_3 = \{\ 2,\quad 4,\ 5\}$

$T_4 = \{\ 2,\quad 4,\ 5\}$

$T_5 = \{1,\quad 3,\quad 5\}$

| $C_3$ | |
|---|---|
| {2,4,5} | |
| {3,4,5} | |

| $L_3$ | |
|---|---|
| {2,4,5} | |
| {3,4,5} | |

| $L_2$ | |
|---|---|
| {2,4} | |
| {2,5} | |
| {3,4} | |
| {3,5} | |
| {4,5} | |

| $L_1$ | |
|---|---|
| {2} | |
| {3} | |
| {4} | |
| {5} | |

- How to generate the candidate list $C_3$ of 3-itemsets?
  1. Join 3 items from $L_1$? E.g., join $\{2\}, \{3\}, \{4\} \in L_1$? Y/N
  2. Join 1 itemset from $L_2$ and 1 item from $L_1$? E.g., join $\{2,4\} \in L_2$ with $\{3\} \in L_1$? Y/N
  3. Join any 2 itemsets from $L_2$? E.g., join $\{2,4\}, \{3,5\} \in L_2$? Y/N
- J______ **step:** Join two frequent $(k-1)$-itemsets with the sets of first $k-2$ items identical.

Why?

# Compute frequent 4-itemsets

$T_1 = \{ \quad 3, \ 4, \ 5\}$

$T_2 = \{ \quad 3, \ 4, \ 5\}$

$T_3 = \{ \ 2, \quad 4, \ 5\}$

$T_4 = \{ \ 2, \quad 4, \ 5\}$

$T_5 = \{1, \ 3, \quad 5\}$

$L_4 = \emptyset$

$L_3$

| | |
|---|---|
| {2,4,5} | |
| {3,4,5} | |

$L_2$

| | |
|---|---|
| {2,4} | |
| {2,5} | |
| {3,4} | |
| {3,5} | |
| {4,5} | |

$L_1$

| | |
|---|---|
| {2} | |
| {3} | |
| {4} | |
| {5} | |

- **Join step:** Join two frequent $(k - 1)$-itemsets with the sets of first $k - 2$ items identical.
- $C_4$ and therefore $L_4$ are empty because ______________________________________________.
- Any more tricks to shorten candidate lists?

# Compute frequent 1-itemsets

$T_1 = \{1, 2, 3\}$

$T_2 = \{1, 2, 3\}$

$T_3 = \{1, 2, 4\}$

$T_4 = \{1, 2, 4\}$

$T_5 = \{1, 3, 5\}$

$C_1$

| | |
|---|---|
| $\{1\}$ | 5 |
| $\{2\}$ | 4 |
| $\{3\}$ | 3 |
| $\{4\}$ | 2 |
| ~~$\{5\}$~~ | ~~1~~ |

$L_1$

| | |
|---|---|
| $\{1\}$ | 5 |
| $\{2\}$ | 4 |
| $\{3\}$ | 3 |
| $\{4\}$ | 2 |

- Rename item $i$ as $6 - i$ for the previous example.

# Compute frequent 2-itemsets

$T_1 = \{1, 2, 3\}$

$T_2 = \{1, 2, 3\}$

$T_3 = \{1, 2, 4\}$

$T_4 = \{1, 2, 4\}$

$T_5 = \{1, 3, 5\}$

$C_2$

| | |
|---|---|
| {1,2} | 4 |
| {1,3} | 3 |
| {1,4} | 2 |
| {2,3} | 2 |
| {2,4} | 2 |
| ~~{3,4}~~ | ~~0~~ |

$L_2$

| | |
|---|---|
| {1,2} | 4 |
| {1,3} | 3 |
| {1,4} | 2 |
| {2,3} | 2 |
| {2,4} | 2 |

$L_1$

| | |
|---|---|
| {1} | 5 |
| {2} | 4 |
| {3} | 3 |
| {4} | 2 |

- **Join step:** Join two frequent $(k-1)$-itemsets with the sets of first $k-2$ items identical.

# Compute frequent 3-itemsets

$T_1 = \{1, 2, 3\}$

$T_2 = \{1, 2, 3\}$

$T_3 = \{1, 2, 4\}$

$T_4 = \{1, 2, 4\}$

$T_5 = \{1, 3, 5\}$

$C_3$

| | |
|---|---|
| {1,2,3} | 2 |
| {1,2,4} | 2 |
| ~~{1,3,4}~~ | |
| ~~{2,3,4}~~ | |

$L_3$

| | |
|---|---|
| {1,2,3} | 2 |
| {1,2,4} | 2 |

$L_2$

| | |
|---|---|
| {1,2} | 4 |
| {1,3} | 3 |
| {1,4} | 2 |
| {2,3} | 2 |
| {2,4} | 2 |

$L_1$

| | |
|---|---|
| {1} | 5 |
| {2} | 4 |
| {3} | 3 |
| {4} | 2 |

- **Join step:** Join two frequent $(k-1)$-itemsets with the set of first $k-2$ items identical.
- **Prune step:** Remove an itemset from $C_k$ if any of its $(k-1)$-subsets is not in $L_{k-1}$.

# Compute frequent 4-itemsets

$T_1 = \{1, 2, 3\}$

$T_2 = \{1, 2, 3\}$

$T_3 = \{1, 2, 4\}$

$T_4 = \{1, 2, 4\}$

$T_5 = \{1, 3, 5\}$

$C_4 = \emptyset = L_4$

| |
|---|
| $\cancel{\{1,2,3,4\}}$ |

$L_3$

| | |
|---|---|
| {1,2,3} | 2 |
| {1,2,4} | 2 |

$L_2$

| | |
|---|---|
| {1,2} | 4 |
| {1,3} | 3 |
| {1,4} | 2 |
| {2,3} | 2 |
| {2,4} | 2 |

$L_1$

| | |
|---|---|
| {1} | 5 |
| {2} | 4 |
| {3} | 3 |
| {4} | 2 |

- **Join step:** Join two frequent $(k-1)$-itemsets with the sets of first $k-2$ items identical.
- **Prune step:** Remove an itemset from $C_k$ if any of its $(k-1)$-subsets is not in $L_{k-1}$.
  - (Check at most ______________________________ subsets, which are obtained by removing ______________________________________________________________________________.)

# Apriori algorithm

- Compute frequent $k$-itemsets for $k$ from 1 to $m$:
  - **Join step:** Join two frequent $(k-1)$-itemsets with the sets of first $k-2$ items identical.
  - **Prune step:** Remove an itemset from $C_k$ if any of its $(k-1)$-subsets is not in $L_{k-1}$.
- Complexity: ______________________________.
  - $|L_k|$ can go up to ________________________.
  - The total number of (non-empty) frequent itemsets can go up to _______________________.
- Can we compute and store $L$ more efficiently?

# How to store frequent itemsets

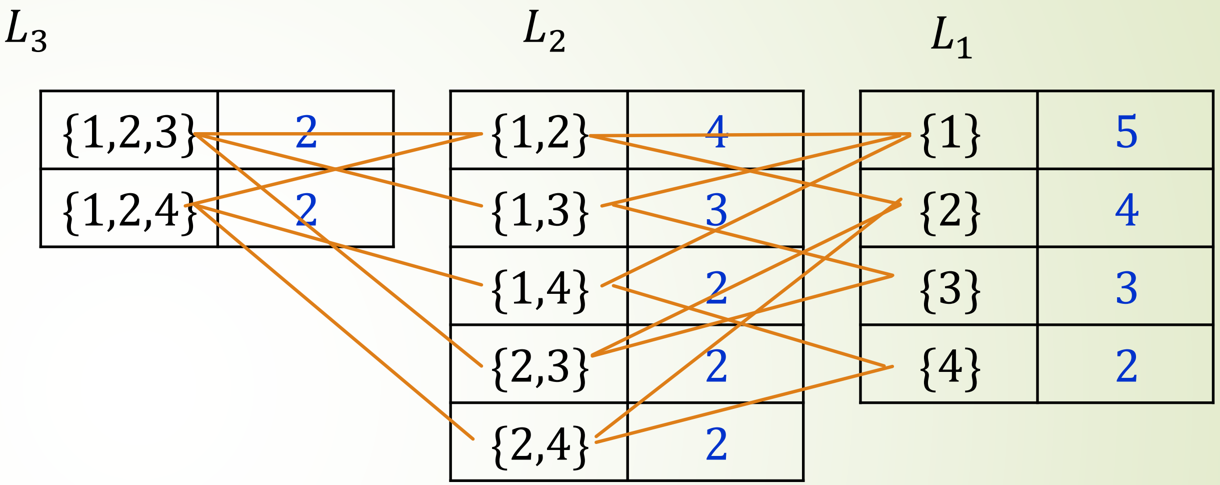
$T_1 = \{1,\ 2,\ 3\}$

$T_2 = \{1,\ 2,\ 3\}$

$T_3 = \{1,\ 2,\ \quad 4\}$

$T_4 = \{1,\ 2,\ \quad 4\}$

$T_5 = \{1,\ \quad 3,\ \quad 5\}$



$L_3$

| | |
|---|---|
| {1,2,3} | 2 |
| {1,2,4} | 2 |

$L_2$

| | |
|---|---|
| {1,2} | 4 |
| {1,3} | 3 |
| {1,4} | 2 |
| {2,3} | 2 |
| {2,4} | 2 |

$L_1$

| | |
|---|---|
| {1} | 5 |
| {2} | 4 |
| {3} | 3 |
| {4} | 2 |

- A frequent itemset is **m**_______________ iff all its proper supersets are not frequent.
- Can store only the maximal frequent itemsets because
  - All non-empty s_____________ of a maximal frequent itemset are frequent.
  - Every frequent itemset is a s______________ of a maximal frequent itemset.
- How to store the counts of the frequent itemsets?

# How to store frequent itemsets with their counts

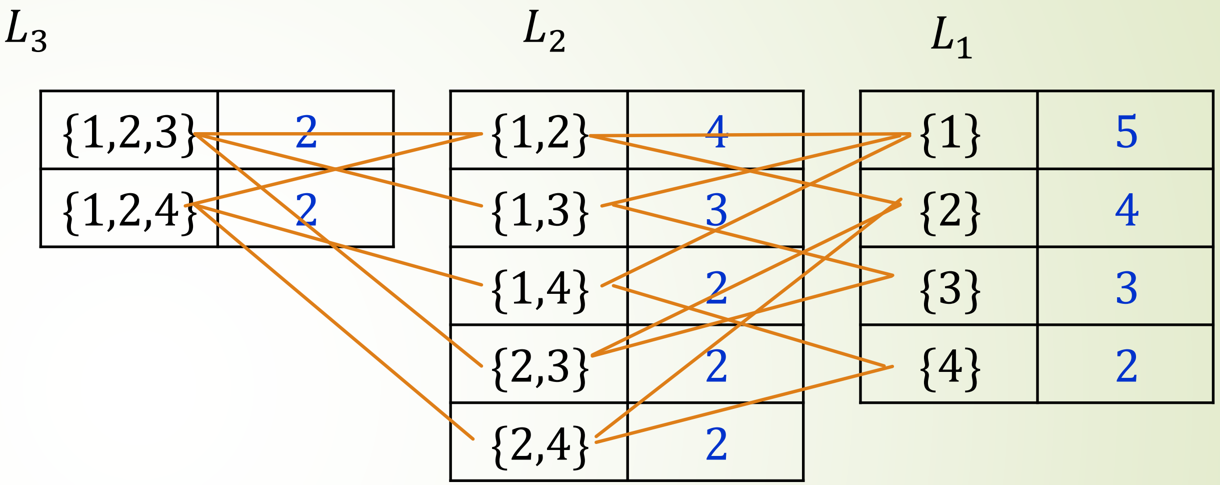
$T_1 = \{1, 2, 3\}$

$T_2 = \{1, 2, 3\}$

$T_3 = \{1, 2, 4\}$

$T_4 = \{1, 2, 4\}$

$T_5 = \{1, 3, 5\}$



$L_3$

| | |
|---|---|
| {1,2,3} | 2 |
| {1,2,4} | 2 |

$L_2$

| | |
|---|---|
| {1,2} | 4 |
| {1,3} | 3 |
| {1,4} | 2 |
| {2,3} | 2 |
| {2,4} | 2 |

$L_1$

| | |
|---|---|
| {1} | 5 |
| {2} | 4 |
| {3} | 3 |
| {4} | 2 |

- An itemset is **c**___________ iff no item can be added without decreasing the count.
- Can store only closed frequent itemsets and their counts because
  - maximal frequent itemsets are closed/not closed, so all frequent itemsets can be recovered.
  - a frequent itemset $A$ has the same count as a closed frequent itemset containing/contained by $A$ with largest/smallest count.
- [Challenge] How about storing itemsets where no item can be removed without increasing the count?

# Computational efficiency (Optional)

- Improving the efficiency of Apriori (6.2.3)
  - Hash-based technique
  - Transaction reduction
  - Partitioning
  - Sampling
  - Dynamic itemset counting
- Mining closed and Max Patterns (6.2.6)
  - Item merging
  - Sub-itemset pruning
  - Item skipping

# Association rules

- If a customer has bought all the items in $A$, is he/she likely to buy items in $B$? Why is it useful to know? **R**________________ **s**___________________
- Association rules:
  - $A \subseteq \mathrm{T} \Rightarrow B \subseteq \mathrm{T}$ for a random transaction $\mathrm{T}$.
  - $A \Rightarrow B$ for short.
- Simplifying assumptions:
  - $A \cap B = \emptyset$: without loss of generality since $A \Rightarrow B$ is equivalent to $A \Rightarrow B \backslash \mathrm{A}$.
  - $B \neq \emptyset$: avoid triviality since $A \Rightarrow \emptyset$ is always __________.
  - $A \neq \emptyset$ may be imposed because $\emptyset \Rightarrow B$ is not an a_________________.

# Example
## Perfect rules

$T_1 = \{1, 2, 3\}$

$T_2 = \{1, 2, 3\}$

$T_3 = \{1, 2, 4\}$

$T_4 = \{1, 2, 4\}$

$T_5 = \{1, 3, 5\}$

- $\{2\} \Rightarrow \{1\}$? ______
- $\{3\} \Rightarrow \{1\}$? ______
- $\{4\} \Rightarrow \{1,2\}$? ______
- $\{3\} \Rightarrow \{1,2\}$? ______
- $\{5\} \Rightarrow \{1,3\}$? ______
- $\{3\} \Rightarrow \{5\}$? ______
- $\{1\} \Rightarrow \{5\}$? ______
- $\emptyset \Rightarrow \{1\}$? ______
- $3 \notin \mathrm{T} \Rightarrow 4 \in \mathrm{T}$? ______

# Measures of rule quality

- How many instances satisfy the antecedent?

$$\mathrm{coverage}(A \Rightarrow B) \coloneqq \frac{\mathrm{count}(A)}{n}$$

- How many instances satisfy both the antecedent and consequence?

$$\mathrm{support}(A \Rightarrow B) \coloneqq \frac{\mathrm{count}(A \cup B)}{n} \approx \Pr(A \cup B \subseteq \mathrm{T})$$

- Out of those instances that satisfy the antecedent, how many satisfy the consequence?

$$\mathrm{confidence}(A \Rightarrow B) \coloneqq \frac{\mathrm{count}(A \cup B)}{\mathrm{count}(A)} = \frac{\mathrm{support}(A \Rightarrow B)}{\mathrm{coverage}(A \Rightarrow B)} \approx \Pr(B \subseteq \mathrm{T} | A \subseteq \mathrm{T})$$

- **Support-confidence framework**: Prefer rules with high support and confidence.

# Example

$T_1 = \{1, 2, 3\}$

$T_2 = \{1, 2, 3\}$

$T_3 = \{1, 2, 4\}$

$T_4 = \{1, 2, 4\}$

$T_5 = \{1, 3, 5\}$

| Rule | Coverage | Support | Confidence |
|---|---|---|---|
| $\{2\} \Rightarrow \{1\}$ | 80% | 80% | 100% |
| $\{3\} \Rightarrow \{1\}$ | 60% | 60% | 100% |
| $\{4\} \Rightarrow \{1,2\}$ | 40% | 40% | 100% |
| $\{3\} \Rightarrow \{1,2\}$ | | | |
| $\{5\} \Rightarrow \{1,3\}$ | | | |
| $\{3\} \Rightarrow \{5\}$ | | | |
| $\{1\} \Rightarrow \{5\}$ | | | |

# Association rules from frequent itemsets

- Goal: Obtain all association rules with support ≥ $s$ and confidence≥ $c$.
- How?

1. Generate the list $L$ of frequent item sets with $\text{min\_sup} = \lceil ns \rceil$. ✓
2. For $C \in L: |C| \geq 2$, find non-empty proper subset $A \subseteq C$ with $\text{count}(A) \leq \text{count}(C)/c$ ✓ to generate the rule

$$\boxed{A \Rightarrow B,} \qquad \text{where } B := C \setminus A$$

$\subseteq C \in L$

- Correctness: ✓
  - $\text{support}(A \Rightarrow B) \geq s$ iff $\text{count}(A \cup B) \geq$ ns
  - $\text{confidence}(A \Rightarrow B) \geq c$ iff $\text{count}(A) \leq$ count(A∪B)/c ✓

$$= \frac{\text{count}(A \cup B)}{\text{count}(A)} \geq c$$

$$\frac{\text{count}(A \cup B)}{n} \geq \frac{\lceil ns \rceil}{n} \geq s$$

# Association rules from frequent itemsets

$T_1 = \{1, 2, 3\}$

$T_2 = \{1, 2, 3\}$

$T_3 = \{1, 2, 4\}$

$T_4 = \{1, 2, 4\}$

$T_5 = \{1, 3, 5\}$

50%

$L_3$

| | |
|---|---|
| {1,2,3} | 2 |
| {1,2,4} | 2 |

$L_2$

| | |
|---|---|
| {1,2} | 4 |
| {1,3} | 3 |
| {1,4} | 2 |
| {2,3} | 2 |
| {2,4} | 2 |

$L_1$

| | |
|---|---|
| {1} | 5 |
| {2} | 4 |
| {3} | 3 |
| {4} | 2 |

$s = 0.4$

- With min_sup=2, can generate all rules with support at least $\frac{2}{5} = 0.4$.
- With $c = 0.6$ and $C = \{1,2,3\} \in L_3$, the desired association rules have $\text{count}(A) \leq \frac{\text{count}(\{1,2,3\})}{0.6}$

$$= \frac{2}{0.6} = 3.\dot{3}$$

$\{1,2\} \Rightarrow \{3\}$ Y/[✓ N] $\{1\} \Rightarrow \{2,3\}$ Y/[✓ N]

$\{1,3\} \Rightarrow \{2\}$ [✓ Y]/N $\{2\} \Rightarrow \{1,3\}$ Y/[✓ N]

$\{2,3\} \Rightarrow \{1\}$ [✓ Y]/N $\{3\} \Rightarrow \{1,2\}$ [✓ Y]/N

- Exercise: Continue for other non-trivial choices of $C$: $\{1,2,4\}, \{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}$.

# Limitation of support-confidence framework

$T_1 = \{1, 2, 3\}$

$T_2 = \{1, 2, 3\}$

$T_3 = \{1, 2, 4\}$

$T_4 = \{1, 2, 4\}$

$T_5 = \{1, 3, 5\}$

4

| # | Rule | Coverage | Support | Confidence |
|---|---|---|---|---|
| 1 | $\{1,2\} \Rightarrow \{3\}$ | 80% | 40% | ~~50%~~ |
| 2 | $\{1,3\} \Rightarrow \{2\}$ | 60% | 40% | $6\dot{6}\%$ |
| 3 | $\{2,3\} \Rightarrow \{1\}$ | 40% | 40% | 100% |
| 4 | $\{1\} \Rightarrow \{2,3\}$ | 100% | 40% | 40% |
| 5 | $\{2\} \Rightarrow \{1,3\}$ | 80% | 40% | 50% |
| 6 | $\{3\} \Rightarrow \{1,2\}$ | 60% | 40% | $6\dot{6}\%$ |
| ⋮ | | | | |

- Which rule above has the maximum confidence? Rule #__3__
- Is it the best rule that captures the strongest association? Y/[✓ N] because
  1 is purchased regardless of whether 2, 3 are purchased.

# Limitation of support-confidence framework

$T_1 = \{1, 2, 3\}$

$T_2 = \{1, 2, 3\}$

$T_3 = \{1, 2, 4\}$

$T_4 = \{1, 2, 4\}$

$T_5 = \{1, 3, 5\}$

| # | Rule | Coverage | Support | Confidence | Prior | Lift |
|---|---|---|---|---|---|---|
| 1 | $\{1,2\} \Rightarrow \{3\}$ | 80% | 40% | 50% | 60% | $0.8\dot{3}$ |
| 2 | $\{1,3\} \Rightarrow \{2\}$ | 60% | 40% | $6\dot{6}\%$ | 80% | $0.8\dot{3}$ |
| 3 | $\{2,3\} \Rightarrow \{1\}$ | 40% | 40% | 100% | 100% | 1 |
| 4 | $\{1\} \Rightarrow \{2,3\}$ | 100% | 40% | 40% | | |
| 5 | $\{2\} \Rightarrow \{1,3\}$ | 80% | 40% | 50% | | |
| 6 | $\{3\} \Rightarrow \{1,2\}$ | 60% | 40% | $6\dot{6}\%$ | | |
| ⋮ | | | | | | |

- How much more likely for the consequence to happen if the antecedent is satisfied?

$$\text{lift}(A \Rightarrow B) := \frac{\text{confidence}(A \Rightarrow B)}{\text{prior}(A \Rightarrow B)} = \frac{\text{count}(A \cup B) \cdot n}{\text{count}(A)\text{count}(B)}$$

$$\text{prior}(A \Rightarrow B) = \frac{\text{count}(B)}{n}$$

$$\frac{\Pr(B \subseteq T \mid A \subseteq T)}{\Pr(B \subseteq T)} \quad \cancel{\Pr(A \cup B \subseteq T)} \quad =$$

- Lift is 1 if and only if $A \subseteq \mathrm{T}$ and $B \subseteq \mathrm{T}$ are independent.
- Give 2 rule(s) with positive association, i.e. lift>1: $\{5\} \Rightarrow \{1, 3\}$ with lift $= \frac{1}{0.6} = 1.\dot{6}$

# References

- 6.2.1 Apriori Algorithm: Finding Frequent Itemsets by Confined Candidate Generation
- 6.2.2 Generating Association Rules from Frequent itemsets
- Optional:
  - 6.2.2 Generating Association Rules from Frequent itemsets
  - 6.2.3 Improving the Efficiency of Apriori
  - 6.2.6 Mining closed and Max Patterns
  - Azevedo, Paulo J., and Alípio M. Jorge. "Comparing rule measures for predictive association rules." *European Conference on Machine Learning*. Springer, Berlin, Heidelberg, 2007.