Outline

• Motivation and applications
• Itemsets
• The itemset lattice and the downward closure
• Association rules
You run an on-line store, and want to increase sales.

**Associative advertising**: show ads of relevant products *before* your users search for these

Need the sets of products that are *frequently* bought together
Key definitions

- Items
- Dataset
- Transaction
- Itemset
- Support / frequency,
- Minimum support frequency threshold,
- Frequent itemsets
# Market basket data

- **Items for sale:** \( I = \{\text{apple, beer, cola, diapers, eggs}\} \)
- **Transactions:**
  1: \{apple, cola\},
  2: \{apple, beer, diapers, eggs\},
  3: \{cola, beer, diapers\},
  4: \{apple, beer, cola, diapers\},
  5: \{apple, cola, diapers\}

<table>
<thead>
<tr>
<th>TID</th>
<th>Apple</th>
<th>Beer</th>
<th>Cola</th>
<th>Diapers</th>
<th>Eggs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>2</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>3</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>4</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>6</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Transaction Data as a Binary Matrix**

<table>
<thead>
<tr>
<th>TID</th>
<th>Apple</th>
<th>Beer</th>
<th>Cola</th>
<th>Diapers</th>
<th>Eggs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Any** data that can be represented as a binary matrix can be used.
Applications –

- **Baskets** = sentences; **Items** = documents containing those sentences
  - Items that appear together too often could represent plagiarism
  - Items do not have to be “in” baskets

- **Baskets** = patients; **Items** = drugs & side-effects
  - Has been used to detect combinations of drugs that result in particular side-effects
  - **Requires extension:** Absence of an item needs to be observed as well as presence
Example: Frequent Itemsets

- **Items** = \{milk, coke, pepsi, beer, juice\}
- **Support threshold** = 3 baskets
  - \(B_1 = \{m, c, b\}\)
  - \(B_2 = \{m, p, j\}\)
  - \(B_3 = \{m, b\}\)
  - \(B_4 = \{c, j\}\)
  - \(B_5 = \{m, p, b\}\)
  - \(B_6 = \{m, c, b, j\}\)
  - \(B_7 = \{c, b, j\}\)
  - \(B_8 = \{b, c\}\)

- **Frequent itemsets**: \{m\}, \{c\}, \{b\}, \{j\}, \{m,b\}, \{b,c\}, \{c,j\}
Search space and search strategy
The Itemset Lattice

\{apple, cola, dog food\} \rightarrow \{beer, cola, dog food\}

\{apple, cola\} 

\{a, apple\} \rightarrow \{b, beer\} 

\{a, apple\} \rightarrow \{c, cola\} 

\{a, apple\} \rightarrow \{d, dog food\}

\{a, apple\} 

\{a\, apple\} 

\emptyset
Naïve search strategy for FIs

Try every possible itemset and check if it is frequent

Qs:

• How to try every itemset?
• How to check if it is frequent? How long does it take?
• How long does it take to mine all fis?
The Itemset Lattice

data

itemset lattice
The Itemset Lattice

data

itemset lattice
Downward closure property

a.k.a. Antimonotonicity of support