# Constraint based

T.R.Lekhaa
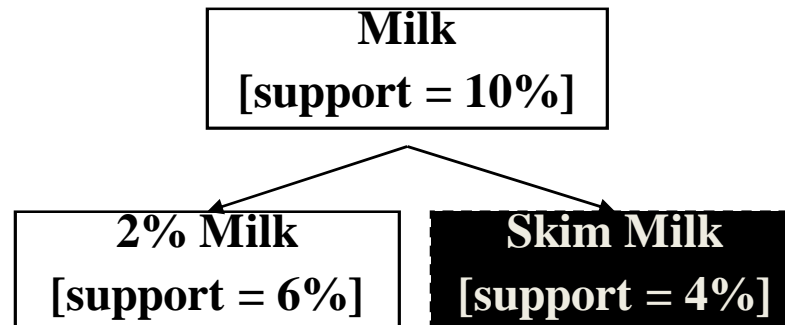
AP/IT

# Mining Multiple-Level Association Rules

- Items often form hierarchies
- Flexible support settings
  - Items at the lower level are expected to have lower support
- Exploration of *shared* multi-level mining

uniform support                                                    reduced support

**Level 1**
**min_sup = 5%**

**Level 1**
**min_sup = 5%**

```
              ┌─────────────────────┐
              │       Milk          │
              │ [support = 10%]     │
              └─────────────────────┘
                    ╱        ╲
   ┌──────────────────┐   ┌──────────────────┐
   │     2% Milk      │   │    Skim Milk     │
   │ [support = 6%]   │   │ [support = 4%]   │
   └──────────────────┘   └──────────────────┘
```

**Level 2**
**min_sup = 5%**

**Level 2**
**min_sup = 3%**

# Multi-level Association: Flexible Support and Redundancy filtering

- Flexible min-support thresholds: Some items are more valuable but less frequent
  - Use non-uniform, group-based min-support
  - E.g., {diamond, watch, camera}: 0.05%; {bread, milk}: 5%; ...
- Redundancy Filtering: Some rules may be redundant due to "ancestor" relationships between items
  - milk $\Rightarrow$ wheat bread  [support = 8%, confidence = 70%]
  - 2% milk $\Rightarrow$ wheat bread [support = 2%, confidence = 72%]

  The first rule is an ancestor of the second rule
- A rule is *redundant* if its support is close to the "expected" value, based on the rule's ancestor

# Constraint-Based Frequent Pattern Mining

- Anti-monotonic: If constraint c is violated, its further mining can be terminated

- Monotonic: If c is satisfied, no need to check c again

- Succinct: c must be satisfied, so one can start with the data sets satisfying c

- Convertible: c is not monotonic nor anti-monotonic, but it can be converted into it if items in the transaction can be properly ordered

# erestingness Measure: Correlations (Lift)

- *play basketball* $\Rightarrow$ *eat cereal* [40%, 66.7%] is misleading

  - The overall % of students eating cereal is 75% > 66.7%.

- *play basketball* $\Rightarrow$ *not eat cereal* [20%, 33.3%] is more accurate, although with lower support and confidence

- Measure of dependent/correlated events: lift

$$lift = \frac{P(A \cup B)}{P(A)P(B)}$$

| | Basketball | Not basketball | Sum (row) |
|---|---|---|---|
| Cereal | 2000 | 1750 | 3750 |
| Not cereal | 1000 | 250 | 1250 |
| Sum(col.) | 3000 | 2000 | 5000 |

$$lift(B,C) = \frac{2000/5000}{3000/5000 * 3750/5000} = 0.89$$

$$lift(B,\neg C) = \frac{1000/5000}{3000/5000 * 1250/5000} = 1.33$$