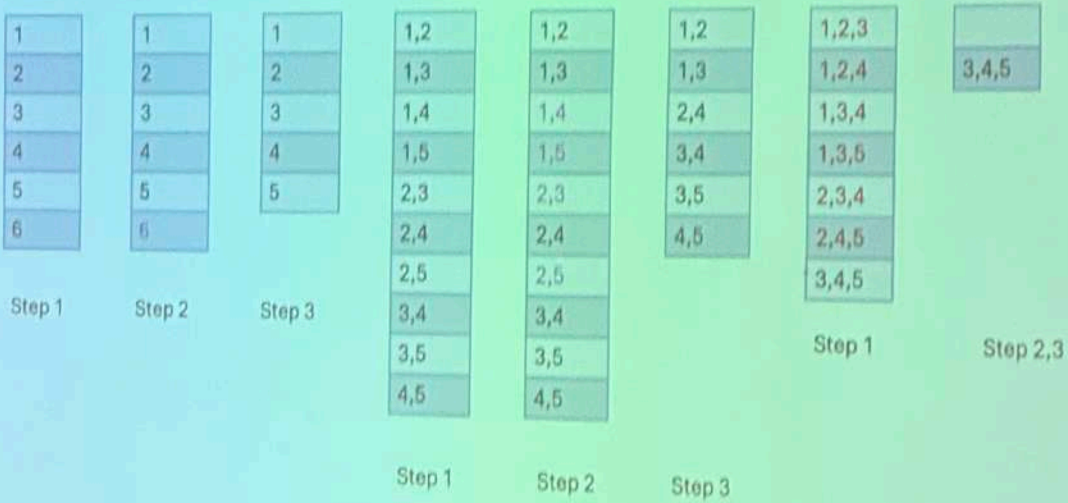
Apriori Algorithm

- Invented by Rakesh Agrawal and Ramakant Srikant (1994)
- Can we speed up than pure brute force?

- Apriori: acknowledges the prior knowledge
  - If any itemset is not frequent, its superset cannot be frequent
  - An itemset can be frequent only if all its subsets are frequent

How does it work?

- Step 0: create 1-size frequent *itemsets* list that meet threshold support, k=1
- Step 1: Expand the *itemsets* list
  - From the k sized *itemsets* list combine overlapping sets to k+1 size *itemsets* list
- Step 2: Prune the expanded *itemsets* list using apriori property,
  - k=k+1
- Step 3: remove infrequent *itemsets* from the list
- Repeat Step 1,2,3 till no more further expansion possible



L_1 $\{1\},\{2\},\{3\},\{4\},\{5\}$, L_2 $\{1,2\},\{1,3\},\{2,4\},\{3,4\},\{3,5\},\{4,5\}$, $L3$ $\{3,4,5\}$

```
Apriori(T, ε)
    L₁ ← {large singleton itemsets}
    k ← 2
    while L_{k-1} is not empty
        C_k ← Generate_candidates(L_{k-1}, k)
        for transactions t in T
            D_t ← {c in C_k : c ⊆ t}
            for candidates c in D_t
                count[c] ← count[c] + 1

        L_k ← {c in C_k : count[c] ≥ ε}
        k ← k + 1

    return Union(L_k) over all k

Generate_candidates(L, k)
    result ← empty_set()
    for all p ∈ L, q ∈ L where p and q differ in exactly one element
        c ← p ∪ q
        if u ∈ L for all u ⊆ c where |u| = k-1
            result.add(c)
    return result
```

Image credits: wikipedia

- Different types of association rules (Categorical, hierarchical, cyclic)

- Eclat Algorithm
  - Equivalence Class Transformation: a depth-first search strategy

- FP-Growth
  - Frequent Pattern Growth: a compact data structure called the FP-tree (Frequent Pattern tree) to compress the dataset

Applications

- Text classification
  - Classify emails into spam / non-spam
  - NLP Problems
    - Tagging: Classify words into verbs, nouns, etc.
- Risk management, Fraud detection, Computer intrusion detection
  - Given the properties of a transaction (items purchased, amount, location, customer profile, etc.)
  - Determine if it is a fraud
- Machine learning / pattern recognition applications
  - Vision
  - Speech recognition etc.
- All of science & knowledge is about predicting future in terms of past
  - So classification is a very fundamental problem with ultra-wide scope of applications

- We collect different measurements/facts/about certain features
- $x = (x_1, x_2, \dots, x_d)$
  - In the above example, x_1, x_2 are diameter and weight
- $y \in \{1,2,\dots,K\} = [K]$
- If $K = 2$ binary classification, else, multi-class classification

- What is classifier?
- How to measure 'goodness' of a classifier?
- If only 1% population has cancer, then a test for cancer that classifies all people as non-cancer will predict 99% of the trails correctly

$y_i \backslash \hat{y}_i$	0	1
0	TN	FP
1	FN	TP

$P = FN + TP$

$N = TN + FP$

$$\text{Accuracy} = \frac{TP + TN}{P + N}$$

$$\text{Recall / sensitivity} = \frac{TP}{P} = \frac{TP}{TP + FN} \quad (TPR)$$

$$\text{Precision} = \frac{TP}{FP + TP}$$

(Positive predicti
value
PPV)

$$FNR = \frac{FN}{P} \qquad TPR + FNR = 1$$

$$TNR = \frac{TN}{N} \qquad TNR + FPR = 1$$

$$FPR = \frac{FP}{N}$$

y_i \ $\hat{y}_i$	0	1
0	TN	FP
1	FN	TP

$$P = FN + TP$$

$$N = TN + FP.$$