# Challenges

- A major challenge in mining frequent itemsets from a large data set is the fact that such mining often generates a huge number of itemsets satisfying the minimum support threshold, especially when *minsup* is set low.
- This is because if an itemset is frequent, each of its subsets is frequent as well.
- A long itemset will contain a combinatorial number of shorter, frequent sub-itemsets.
- For example, a frequent itemset of length 100, such as {a1, a2, , , , a100}, contains $\binom{100}{1}$ =100 frequent 1-itemsets: a1, a2, , , , , a100, $\binom{100}{2}$ frequent 2-itemsets: (a1, a2), (a1, a3), , , , , (a99, a100), and so on. The total number of frequent itemsets that it contains is thus,

$$\binom{100}{1} + \binom{100}{2} + .... + \binom{100}{100} = 2^{100} - 1 \approx 1.27X10^{30} \qquad (3)$$

# Frequent Closed Itemset Mining

## Definition 6 (Frequent Closed Itemset)

An itemset $X$ is called frequent closed itemset if and only if it is frequent and there exists no proper superset $X''$, $(X \subset X'')$ such that support of $X$ is same as the support of $X''$, $sup(X)=sup(X'')$.

## Definition 7 (Frequent Maximal Itemset)

An itemset $X$ is called frequent maximal itemset if and only if it is frequent and there exists no proper superset $X''$, $(X \subset X'')$.

- Suppose that a transaction database has only two transactions: $\{\langle$a1, a2, , , , , a100$\rangle$; $\langle$a1, a2, : : : , a50$\rangle\}$. Let the minimum support count threshold be $minsup = 1$.
- Ttwo closed frequent itemsets and their support counts, that is, C = {{a1, a2, : : : , a100} : 1; {a1, a2, : : : , a50g : 2}}.

# Frequent Closed Itemset Mining

- Suppose that a transaction database has only two transactions: $\{\langle$a1, a2, , , , , a100$\rangle$; $\langle$a1, a2, : : : , a50$\rangle\}$. Let the minimum support count threshold be $minsup = 1$.
- Two closed frequent itemsets and their support counts, that is, C = $\{\{$a1, a2, : : : , a100$\}$ : 1; $\{$a1, a2, , , , , a50g : 2$\}\}$.
- There is one maximal frequent itemset M = $\{\{$a1, a2, , , , , a100$\}$ : 1$\}$. (We cannot include $\{$a1, a2, , , , , a50$\}$ as a maximal frequent itemset because it has a frequent super-set, $\{$a1, a2, , , , , a100$\}$.)

# Frequent Closed Itemset Mining from High Dimensional Dataset

- The conventional algorithms mine frequent itemsets, frequent closed itemset and frequent maximal itemset from the transactional datasets.
- In the modern era, the abundant data across variety of domains, including bioinformatics has led to the new form of dataset known as a high dimensional dataset, whose data characteristics are different from that of transactional datasets.
- The high dimensional datasets consist of less number of rows and considerably large number of features.
- The amount of information that can be extracted from high dimensional datasets is potentially huge, but extraction of information and knowledge from these datasets is a non-trivial task.

# Frequent Closed Itemset Mining from High Dimensional Dataset

- The conventional algorithms adopt feature enumeration based approach for mine frequent closed itemsets.
- The conventional algorithms face an uphill task in mining frequent closed itemsets from the high dimensional datasets.
- To overcome the inefficiency and uphill task of these algorithms, sequential row enumerated algorithms were proposed to mine FCI from high dimensional datasets.
- This problem of inefficiency can be solved to the greater extent by parallel row enumerated algorithms.

# Frequent Colossal Itemset Mining

- The result of frequent closed itemset mining algorithms includes small and mid-sized itemsets, which does not enclose valuable and complete information in many applications.
- In application dealing with high dimensional datasets such as bioinformatics, association rule mining gives greater importance to the large-sized itemsets called as colossal itemsets.

**Definition 8 (Frequent Colossal Itemset)**

An itemset $X$ is called frequent colossal itemset if and only if it is frequent and $card(X) \geq mincard$, where $mincard$ is user specified least cardinality threshold.

**Example 6**

In Table 1, the itemset $X = \{f_1, f_2, f_6, f_{10}\}$ is frequent colossal itemset with minimum support threshold set to 2 and minimum cardinality threshold set to 4, $sup(X) \geq 2$ and $card(X) \geq 4$.

# Frequent Colossal Closed Itemset Mining

**Definition 9 (Frequent Colossal Closed Itemset)**

An itemset $X$ is called frequent colossal closed itemset if and only if it is frequent closed and $card(X) \geq mincard$, where $mincard$ is user specified least cardinality threshold.

**Example 7**

In Table 1, the itemset $X = \{f_2, f_4, f_7, f_8\}$, is frequent colossal closed itemset with minimum support threshold set to 2 and minimum cardinality threshold set to 4, $sup(X) \geq 2$ and $card(X) \geq 4$.