

Apriori Algorithm

Suhas P
1GA19CS159

Apriori Algorithm

- Apriori algorithm is used for frequent itemset mining and association rule mining.
- The algorithm simply follows a basis that any subset of a large itemset must be a large itemset.
- The Apriori principle can reduce the number of itemsets needed to be examined.
- Apriori principle suggests if an itemset is frequent, then all of its subsets must also be frequent.

Assume X and Y are two itemsets. Apriori principle holds due to the following property of support measure:

$$\forall X, Y: (X \subseteq Y) \rightarrow s(X) \geq s(Y)$$

Apriori Algorithm

Apriori algorithm evaluates candidates for association as follows:

C_k : Set of candidate-itemsets of size k

F_l : Set of frequent itemsets of size k

$F_1 = \{\text{large items}\}$

for ($k=1$; $F_k \neq 0$; $k++$) *do* {

C_{k+1} = New candidates generated from F_k

for each transaction t in the database *do*

Increment the count of all candidates in C_{k+1} that are contained in F_{k+1} =
Candidates in C_{k+1} with minimum support }

Apriori Example

TID	Items
1	{A, C, D}
2	{A, B, C, E}
3	{B, E}
4	{B, C, E}

Database



Itemset	Support
{A}	2
{B}	3
{C}	3
{E}	3

Iteration 1: Candidate 1 Itemset



Itemset	Support
{A, B}	1
{A, C}*	2
{A, E}	1
{B, C}*	2
{B, E}*	3
{C, E}*	2

Iteration 2: Candidate 2 Itemset



Subset of a frequent itemset is also frequent

Itemset	Support
{B, C, E}*	2

Iteration 3: Candidate 3 Itemset