

```
import numpy as np

# ts = transections
def load_ts(path_to_data,order):
    ts = []
    with open(path_to_data,'r') as f:
        for lines in f:
            str_line = list(lines.strip().split(','))
            t = list(np.unique(str_line))
            t.sort(key = lambda x: order.index(x))
            ts.append(t)
    return ts

# COUNTING THE OCCURENCES OF EACH ITEM
def count_occurences(itemset,ts):
    count = 0
    for i in range(len(ts)):
        if set(itemset).issubset(set(ts[i])):
            count += 1
    return count

# JOINING TWO ITEM SETS
def join_two_itemsets(it1,it2,order):
    it1.sort(key = lambda x: order.index(x))
    it2.sort(key = lambda x: order.index(x))

    for i in range(len(it1)-1):
        if it1[i] != it2[i]:
            return []

    if order.index(it1[-1]) < order.index(it2[-1]):
        return it1 + [it2[-1]]

    return []

# JOINING THE SET ITEM SETS
def join_set_itemsets(set_of_its, order):
    C = []
    for i in range(len(set_of_its)):
        for j in range(i+1,len(set_of_its)):
            it_out = join_two_itemsets(set_of_its[i],set_of_its[j],order)

            if len(it_out)>0:
                C.append(it_out)
    return C

# GETTING THE FREQUENCY OF ITEMS
def get_frequent(itemsets,tss,min_support,prev_discarded):
    L = []
    supp_count = []
    new_discarded = []
    #num_trans = len(transection)
    k = len(prev_discarded.keys())
```

```

for s in range(len(itemsets)):
    discarded_before = False
    if k>0 :
        for it in prev_discarded[k]:
            if set(it).issubset(set(itemsets[s])):
                discarded_before = True

    if not discarded_before:
        count = count_occurences(itemsets[s],tss)
        if count/len(ts) >= min_support:
            L.append(itemsets[s])
            supp_count.append(count)
        else:
            new_discarded.append(itemsets[s])

return L, supp_count, new_discarded

# MODEL TO PRINT THE TABLE
def print_table(T,supp_count):
    print("Itemset | Frequency")
    for k in range(len(T)):
        print('{ } : { }'.format(T[k],supp_count[k]))
    print()

# GET THE FILE AS INPUT
path_to_data = "apriori_data.txt"
order = ['I' + str(i) for i in range(1,6)]

```

TAKING USER INPUT OF MINIMUM SUPPORT

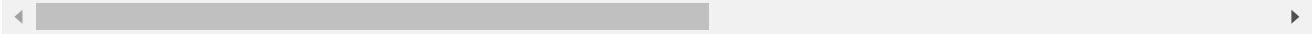
```

min_support = float(input("Enter minimum support(min_sup): "))
print("Minimum support is",min_support)

print(order)
ts = load_ts(path_to_data,order)
print(ts)

Enter minimum support(min_sup): 0.22
Minimum support is 0.22
['I1', 'I2', 'I3', 'I4', 'I5']
[['I1', 'I2', 'I5'], ['I2', 'I4'], ['I2', 'I3'], ['I1', 'I2', 'I4'], ['I1', 'I3'], [

```



```

C = {}
L = {}
itemset_size = 1
discarded = {itemset_size : []}
C.update({itemset_size : [[f] for f in order]})
print(C)

{1: [['I1'], ['I2'], ['I3'], ['I4'], ['I5']]}

supp_count, L = f1

```

```

supp_count_L = {}
f,sup,new_discarded = get_frequent(C[itemset_size],ts,min_support,discarded)
discarded.update({itemset_size : new_discarded})
L.update({itemset_size : f})
supp_count_L.update({itemset_size : sup})

```

PRINTING TABLE L1 AND C1

```

print("Table L1")
print_table(L[1],supp_count_L[1])
print("Table C1")
print_table(C[1],[count_occurences(it,ts) for it in C[1]])

```

Table L1

Itemset	Frequency
['I1']	: 6

['I2']	: 7
--------	-----

['I3']	: 6
--------	-----

['I4']	: 2
--------	-----

['I5']	: 2
--------	-----

Table C1

Itemset	Frequency
['I1']	: 6

['I2']	: 7
--------	-----

['I3']	: 6
--------	-----

['I4']	: 2
--------	-----

['I5']	: 2
--------	-----

```

k = itemset_size + 1
convergence = 0
while convergence == 0:
    C.update({k : join_set_itemsets(L[k-1],order)})
    print("Table C{}:".format(k))
    print_table(C[k],[count_occurences(it,ts) for it in C[k]])
    f, sup, new_discarded = get_frequent(C[k], ts, min_support, discarded)
    discarded.update({k : new_discarded})
    L.update({k : f})
    supp_count_L.update({k : sup})
    if len(L[k]) == 0:
        convergence = True
    else:
        print("Table L{}:".format(k))
        print_table(L[k],supp_count_L[k])
    k += 1

```

Table C2:

```

↳ Itemset | Frequency
['I1', 'I2'] : 4

['I1', 'I3'] : 4

['I1', 'I4'] : 1

['I1', 'I5'] : 2

['I2', 'I3'] : 4

['I2', 'I4'] : 2

['I2', 'I5'] : 2

['I3', 'I4'] : 0

['I3', 'I5'] : 1

['I4', 'I5'] : 0

```

Table L2:

```

Itemset | Frequency
['I1', 'I2'] : 4

['I1', 'I3'] : 4

['I1', 'I5'] : 2

['I2', 'I3'] : 4

['I2', 'I4'] : 2

['I2', 'I5'] : 2

```

Table C3:

```

Itemset | Frequency
['I1', 'I2', 'I3'] : 2

['I1', 'I2', 'I5'] : 2

['I1', 'I3', 'I5'] : 1

['I2', 'I3', 'I4'] : 0

['I2', 'I3', 'I5'] : 1

['I2', 'I4', 'I5'] : 0

```

Table L3:

```

Itemset | Frequency
['I1', 'I2', 'I3'] : 2

['I1', 'I2', 'I5'] : 2

```

Table C4:

```

Itemset | Frequency
['I1', 'I2', 'I3', 'I5'] : 1

```

✓ 0s completed at 22:56 ● ✕