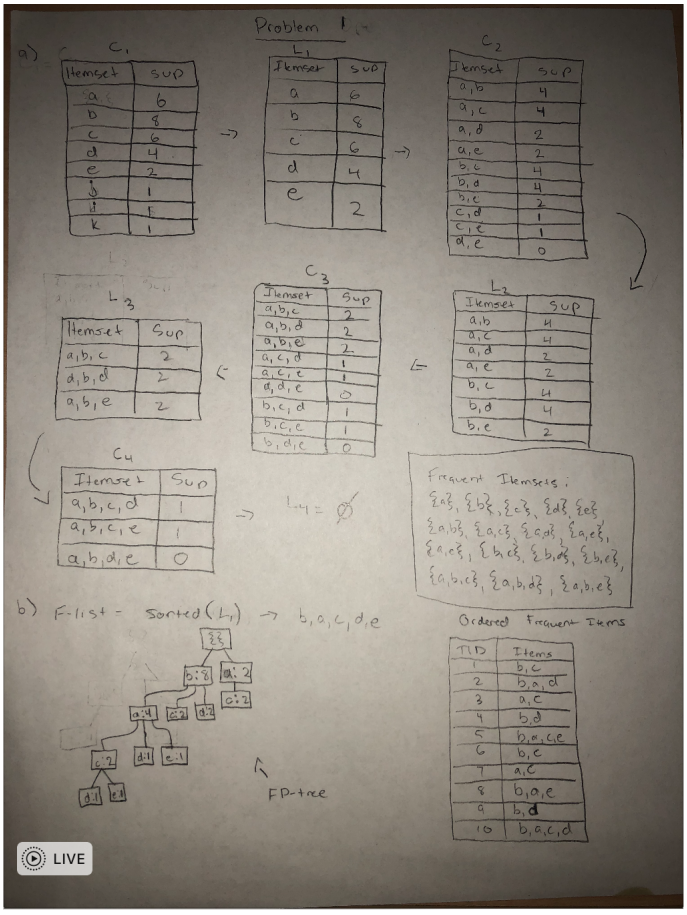
Homework 4 Report

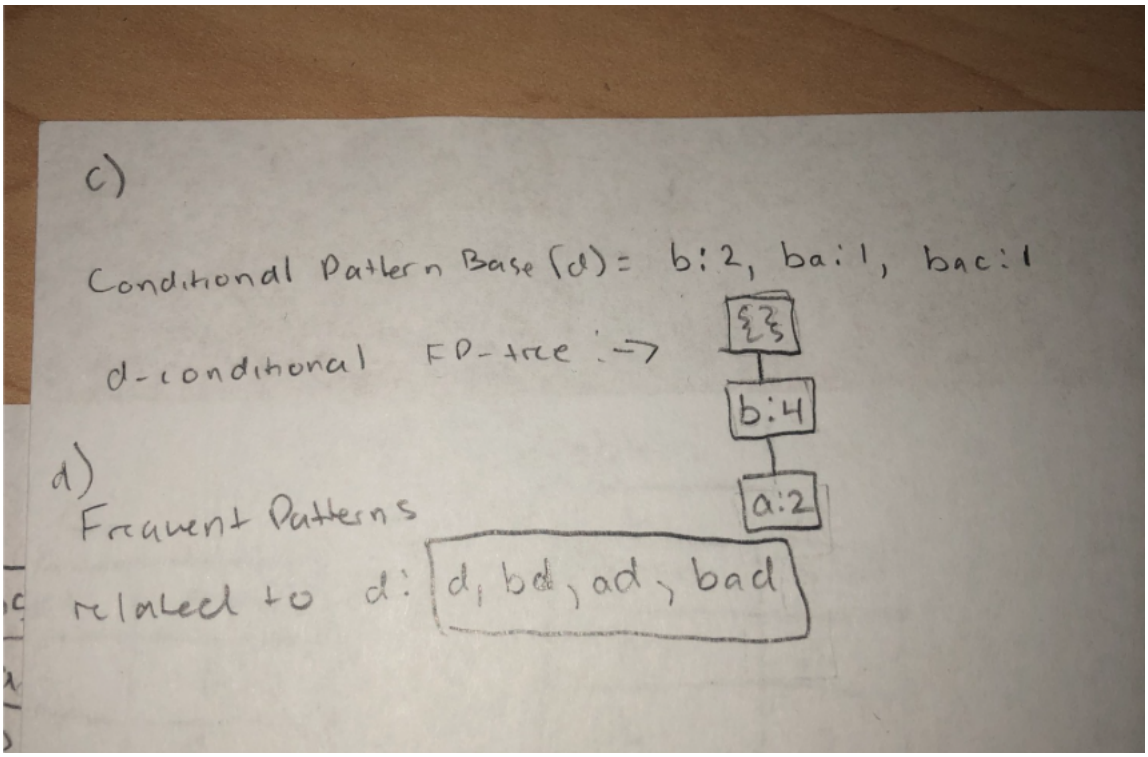
Rajiv Anisetti, UID: 904801422

November 18, 2018

1 Frequent Pattern Mining for Set Data

I did this portion by hand, and have attached pictures of my solution.





2 Apriori for Yelp

The output of the apriori algorithm is as follows:

min_support: 50 min_conf: 0.25
item: "Wicked Spoon","Holsteins Shakes Buns" , 51.000
item: "Wicked Spoon","Secret Pizza" , 52.000
item: "Wicked Spoon","Earl of Sandwich" , 52.000
item: "The Cosmopolitan of Las Vegas","Wicked Spoon" , 54.000
item: "Mon Ami Gabi","Wicked Spoon" , 57.000
item: "Bacchanal Buffet","Wicked Spoon" , 63.000

———————— RULES:
Rule: "Secret Pizza" → "Wicked Spoon" , 0.256
Rule: "The Cosmopolitan of Las Vegas" → "Wicked Spoon" , 0.277
Rule: "Holsteins Shakes Buns" → "Wicked Spoon" , 0.315
263.19302702 sec

These are all Las Vegas locations. This is intuitive; items in frequent itemsets would naturally be close. Wicked Spoon is a very popular buffet, so it has high

frequency. According to the rules, yelping these other locations means there is a degree of confidence they also yelped WS.

3 Correlation Analysis

3.1 Part A

3.1.1 Confidence

Buying Beer $\rightarrow$ Buying Nuts = P(Purchased Nuts and Beer — Purchased Beer) = $\frac{150}{500}$ = **0.3**

Buying Nuts $\rightarrow$ Buying Beer = P(Purchased Beer and Nuts — Purchased Nuts) = $\frac{150}{850}$ = **0.176**

3.1.2 Lift

Lift(Beer, Nuts) = $\frac{P(Buying\ Beer\ and\ Buying\ Nuts)}{P(Buying\ Beer)*P(Buying\ Nuts)}$ = $\frac{\frac{150}{10000}}{\frac{500}{10000}*\frac{850}{10000}}$ = **3.53**

3.1.3 All-confidence

All_Confidence = min(C(Buying Beer $\rightarrow$ Buying Nuts), C(Buying Nuts $\rightarrow$ Buying Beer)) = **0.176**

3.2 Part B

Based on these values, we have a lift > 1, so there is a positive correlation between buying beer and buying nuts. However, there is a higher probability of buying nuts given a beer purchase than the vice versa.

4 GSP Algorithm

4.1 Part A

S contains **4 elements**, and it has **length 6**.

To count subsequences, we must look at the choices for our elements.

We have 2 choices for including a, 2 choices for including b.

There 4 choices for (cd) $\rightarrow$ c, d, (cd), or neither c nor d

The same goes for (ef), so we have $2 * 2 * 4 * 4 = 64$ **combinations total, 63 when not counting the empty subset.**

4.2 Part B

4.2.1 Joining

We can join $< b(cd) >$ and $< (ab)c >$ to form $< (ab)(cd) >$
We can also join $< bce >$ and $< (ab)c$ to form $< (ab)ce >$

4.3 Pruning

Because $< (ab)e >$ cannot be find within L_3, we must prune $< (ab)ce >$

Our overall result for L_4 is $< (ab)(cd) >$