

# Pattern Recognition

## Homework Assignment 1

```
1. H =lil_matrix((N_baskets,M_items), dtype=np.bool)
    for i in range(0,len(data)-1):
        for j in list(map(int,data[i])):
            H[i,j-1] = True
```

The code `range(0,len(data)-1)` does not include the last dataset, to fix it, just remove `-1`. Fixed version is below:

```
H =lil_matrix((N_baskets,M_items), dtype=np.bool)
for i in range(0,len(data)):
    for j in list(map(int,data[i])):
        H[i,j-1] = True
```

2. All rules are in `all_rules.txt`, sorted by confidence from high to low.

The parameters for rules finding:

```
rules = apriori(data, supp=10, zmin=2, zmax=5, target='r',
report='SCL')
```

Note: some of the rules found has lower support than 10%, if you increase the `supp` parameter to 13, then all rules has supports  $\geq 10\%$ .

The top 10 rules :

```
(34, (12, 10, 36, 25), 11.038394415357766, 81.26338329764454,
3.9628866918766232)
(26, (14, 5, 2, 24), 11.227457824316463, 82.12765957446808,
3.24173241810587)
(26, (14, 5, 2, 13), 11.358347876672484, 82.12407991587803,
3.2415911222823035)
(26, (14, 5, 2, 48), 11.896451425247237, 81.96392785571143,
3.2352696207570135)
(26, (14, 5, 2, 28), 11.082024432809773, 81.93548387096774,
3.2341468834487612)
(26, (14, 5, 2, 30), 12.332751599767306, 81.85328185328186,
3.230902215976843)
(26, (14, 5, 2), 12.40546829552065, 81.78331735378715,
3.228140586249371)
(26, (14, 5, 28, 48), 11.28563118091914, 81.4270724029381,
3.2140789313582223)
(26, (14, 5, 13, 24), 10.616637579988366, 81.29175946547885,
3.208737876490428)
(26, (14, 5, 24, 30), 11.72193135543921, 81.25, 3.207089552238806)
```

1. Confidence 81.3, lift 3.96.  
[ education  $\leq$  grade 8, age  $< 34$ , type of house = condominium, dual incomes = not married]  
→ household status = live with parents/family
2. Confidence 82.1, lift 3.24.  
[ occupation = professional/managerial, marital status = married, income  $> \$25000$ , years in Bay Area  $\geq$  four yeas] → dual incomes = yes
3. Confidence 82.1, lift 3.24.  
[ occupation = professional/managerial, marital status = married, income  $> \$25000$ , education  $\geq 1$  to 3 years of college] → dual incomes = yes
4. Confidence 82, lift 3.23.  
[ occupation = professional/managerial, marital status = married, income  $> \$25000$ , language in home = English] → dual incomes = yes
5. Confidence 81.9, lift 3.23.  
[ occupation = professional/managerial, marital status = married, income  $> \$25000$ , number is house household  $\leq 4$ ] → dual incomes = yes
6. Confidence 81.9, lift 3.23.  
[ occupation = professional/managerial, marital status = married, income  $> \$25000$ , number of children  $\leq 4$ ] → dual incomes = yes
7. Confidence 81.8, lift 3.23.  
[ occupation = professional/managerial, marital status = married, income  $> \$25000$ ] → dual incomes = yes
8. Confidence 81.4, lift 3.21.  
[ occupation = professional/managerial, marital status = married, number is house household  $\leq 4$ , language in home = English] → dual incomes = yes
9. Confidence 81.3, lift 3.21.  
[ occupation = professional/managerial, marital status = married, education  $\geq 1$  to 3 years of college, years in Bay Area  $\geq$  four yeas] → dual incomes = yes
10. Confidence 81.3, lift 3.21.  
[ occupation = professional/managerial, marital status = married, years in Bay Area  $\geq$  four yeas, number of children  $\leq 4$ ] → dual incomes = yes