

Lab 2
732A75: Association Analysis -1
Aman Nayak (amana551)
Mahmood Siddique(mahsi404)

Across the experiment we have **SimpleKmean algorithm for Clustering** and **Apriori Algorithm for Association analysis**, advantage of using clustering algorithm is that help in reduction in association analysis time.

Case 1: 3 Clusters and 3 Bins

| |
|-----------------------------|
| Number of Bins: 3 |
| Number of Clusters: 3 |
| lowerBoundMinSupport : 0.05 |
| numRules : 300 |

Best Rule:

Cluster1:

petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 48 ==> cluster=cluster1 48 <conf:(1)>
lift:(2.73) lev:(0.2) [30] conv:(30.4)

Cluster 2:

petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 40 ==> cluster=cluster2 40 <conf:(1)> lift:(3.33)
lev:(0.19) [27] conv:(28)

Cluster 3:

petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 50 ==> cluster=cluster3 50 <conf:(1)> lift:(3)
lev:(0.22) [33] conv:(33.33)

Case 2: 2 Clusters and 3 Bins

| |
|-----------------------------|
| Number of Bins: 3 |
| Number of Clusters: 2 |
| lowerBoundMinSupport : 0.05 |
| numRules : 300 |

Best Rule:

Cluster1:

petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 48 ==> cluster=cluster1 48 <conf:(1)>
lift:(1.56) lev:(0.12) [17] conv:(17.28)

Cluster 2:

sepallength='(5.5-6.7]' petalwidth='(1.7-inf)' 30 ==> cluster=cluster2 30 <conf:(1)> lift:(2.78) lev:(0.13)
[19] conv:(19.2)

Cluster 3:

Null

Case 3: 3 Clusters and 5 Bins

| |
|------------------------------------|
| Number of Bins: 5 |
| Number of Clusters: 3 |
| lowerBoundMinSupport : 0.05 |
| numRules : 300 |

Best Rule:**Cluster1:**

sepalwidth='(2.96-3.44]' petallength='(4.54-5.72]' 25 ==> cluster=cluster1 25 <conf:(1)> lift:(2.38)
lev:(0.1) [14] conv:(14.5)

Cluster 2:

sepalwidth='(2.48-2.96]' petallength='(3.36-4.54]' 18 ==> cluster=cluster2 18 <conf:(1)> lift:(4.29)
lev:(0.09) [13] conv:(13.8)

Cluster 3:

petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' 49 ==> cluster=cluster3 49 <conf:(1)> lift:(2.88)
lev:(0.21) [32] conv:(32.01)

Conclusion:

We have had run a different association algorithm with the different number of clusters and bin combinations since the Apriori Algorithm, which is primarily used here for association analysis, works with discrete data. So we had converted our data in the discrete format using split point between n bins, and we can see that when the number of intervals between discrete data has had been increased using the number of bins, the interval between each attribute had comparatively reduced when compared with the bin as 3 and bin as 5. However, change is not that large, but the same can be gain with a more significant number of bins. Now changed intervals resulted in the different optimal clusters, and thus above results are obtained.