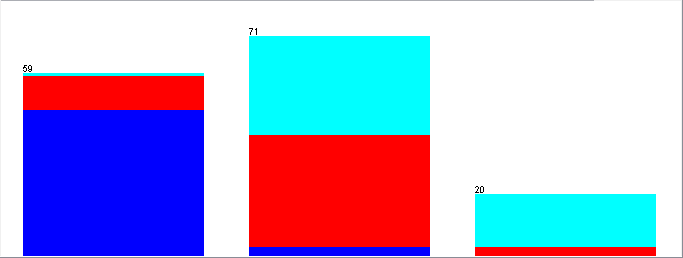
**Lab 2**

**Mohsen Pirmoradiyan, Ahmed Alhasan**

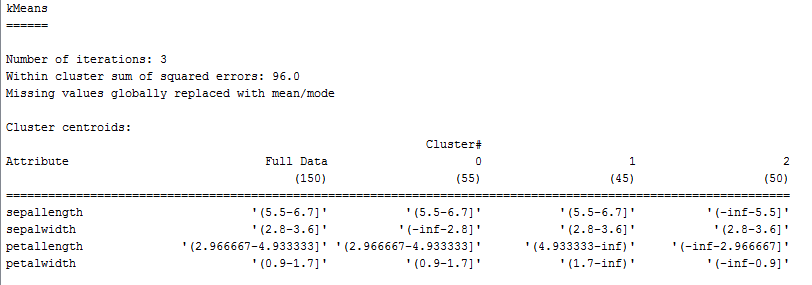
**07 March 2020**

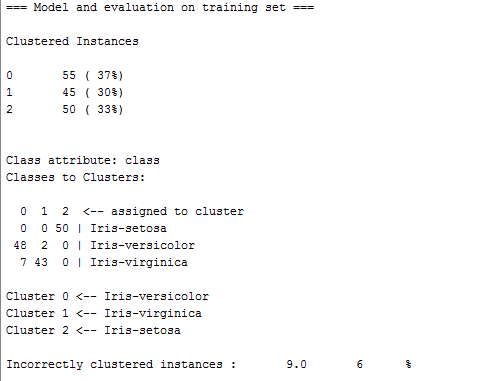
Discretized with 3 bins

****

**K-Means Clustering:**

Clusters = 3, Seed = 10

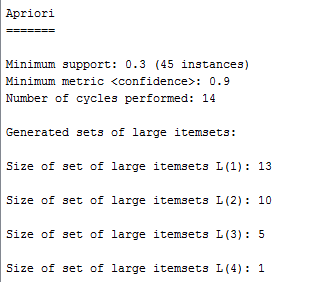


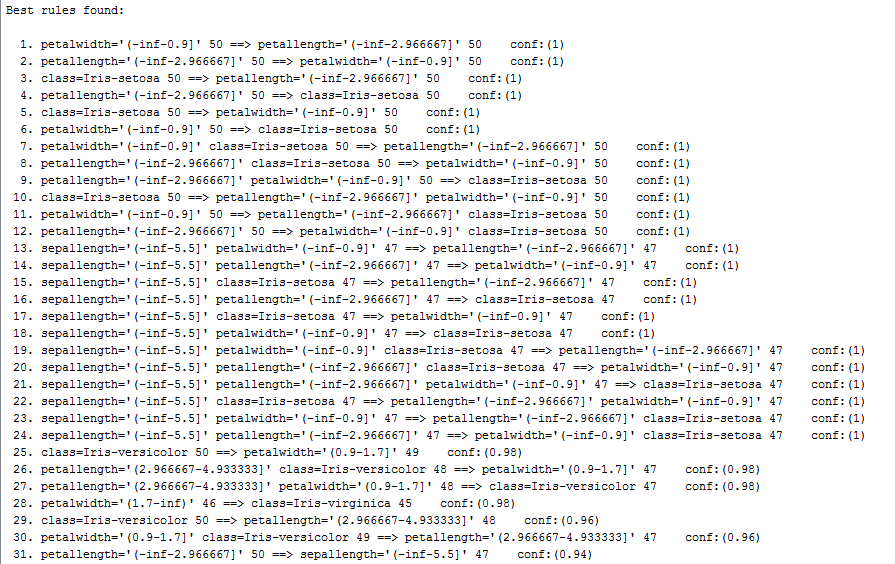


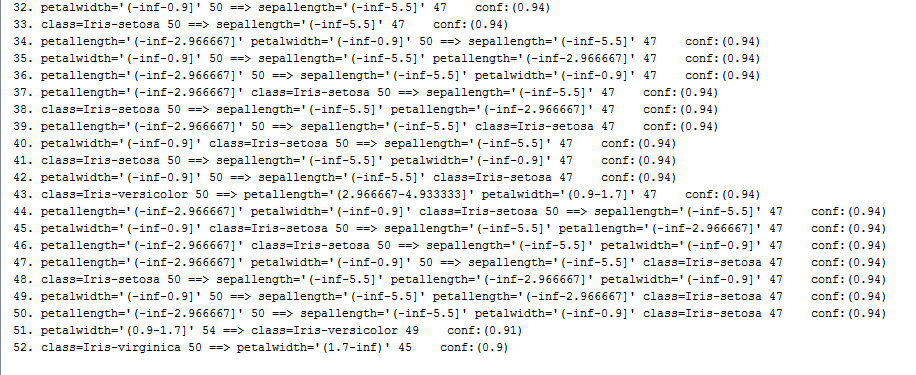
Although K Means is a simple clustering method, it gave good approximation for the 3 types of Iris flowers with only 9 incorrectly clustered instances.

**Association analysis:**

Delta = 0.05, Lower bound of minimum support = 0.3, Confidence Level = 0.9, Number of Rules = 100







In this association analysis we start with upper bound minimum support = 1 (100 %) and decrease by delta = 0.05 at each iteration and stop when either number of rules reach 100 or we get to lower bound minimum support = 0.3.

In this case if we increase lower bound minimum support to 0.4 we will not have any rules generated, meaning there can’t be 60 or more instances out of 150 associated at the same time on two attributes or more.

Also by setting the confidence level to 90% and minimum support to 0.3 we get 52 rules by decreasing any of these parameters we can get much higher number of rules but with less support and/or less confidence level.

**Describing clustering through association analysis**

In all experiments we used seed = 10 for clustering & the apriori algorithm for association. Changes in in parameters from the first experiment are highlighted in yellow.

**Experiment# 1**

Delta = 0.05

Lower bound of minimum support = 0.25

Upper bound of minimum support = 1

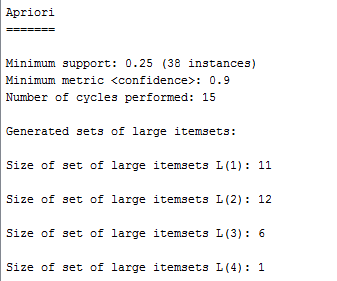
Confidence Level = 0.9

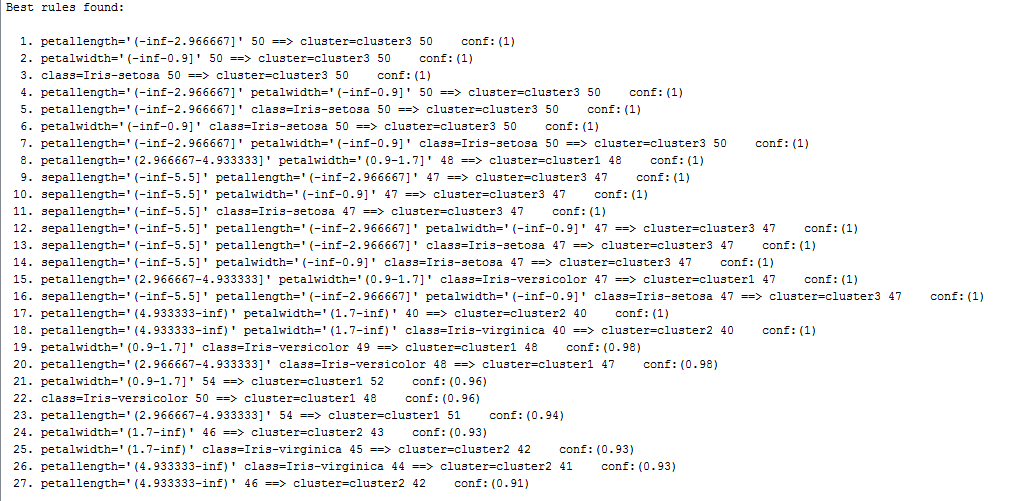
Number of Rules = 100

Clustering Algorithm = K Means Clustering

Clusters = 3

Bins = 3

****



In this experiment with the parameters selected above, we got 27 rules where the best rule for cluster 3 is:

1. petallength='(-inf-2.966667]' 50 ==> cluster=cluster3 50 <conf:(1)>

for cluster 1:

8. petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 48 ==> cluster=cluster1 48 <conf:(1)>

And for cluster 2:

18. petallength='(4.933333-inf)' petalwidth='(1.7-inf)' class=Iris-virginica 40 ==> cluster=cluster2 40 <conf:(1)>

This is because these intervals (bins) are the most predominant ones in each of their respective clusters.

Also because the we only have 3 bins the maximum number of each interval can not exceed 50 and so is the support can not exceed 50/150.

**Experiment# 2**

Delta = 0.05

Lower bound of minimum support = 0.02

Upper bound of minimum support = 1

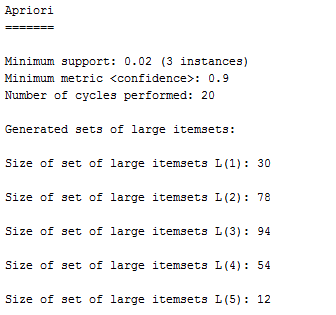
Confidence Level = 0.8

Number of Rules = 200

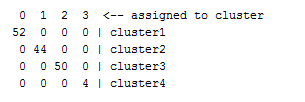
Clustering Algorithm = K Means Clustering

Clusters = 4

Bins = 3



In this experiment we changed the number of clusters to 4, and we had to change the minimum support to 0.02 and number of rules to 200 and that’s to accommodate for the few instances in Cluster 4



The best rule for each cluster is as the following:

**Cluster 1:**

31. sepallength='(5.5-6.7]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 33 ==> cluster=cluster1 33 conf:(1)

**Cluster 2:**

15. petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 40 ==> cluster=cluster2 40 conf:(1)

**Cluster 3:**

1. petallength='(-inf-2.966667]' 50 ==> cluster=cluster3 50 conf:(1)

**Cluster 4:**

114. sepallength='(6.7-inf)' sepalwidth='(2.8-3.6]' petalwidth='(0.9-1.7]' 3 ==> cluster=cluster4 3 conf:(1)

Because we have changed the number of clusters to 4 we got cluster 4 with only 4 instances therefore this cluster can not support rules with more than 4/150.

**Experiment# 3**

Delta = 0.05

Lower bound of minimum support = 0.2

Upper bound of minimum support = 1

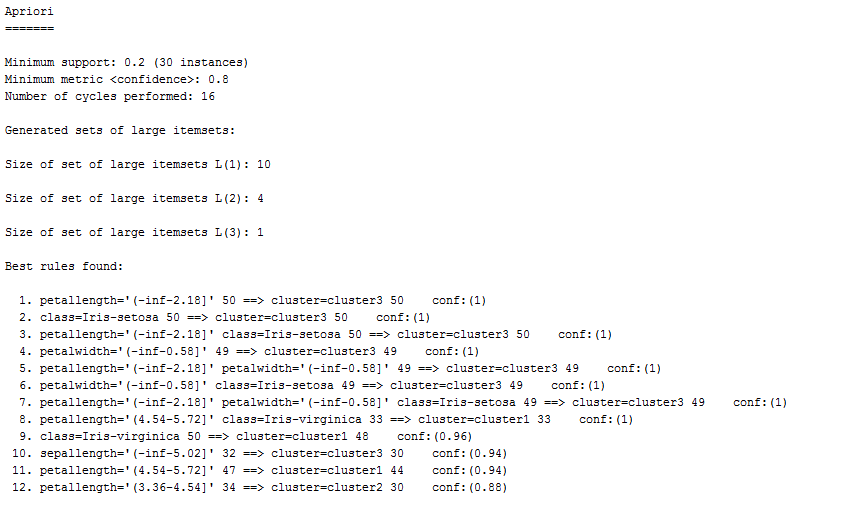
Confidence Level = 0.8

Number of Rules = 100

Clustering Algorithm = K Means Clustering

Clusters = 3

Bins = 5



In this experiment because we changed the number of bins the intervals are now different.

**Cluster 1:**

8. petallength='(4.54-5.72]' class=Iris-virginica 33 ==> cluster=cluster1 33 <conf:(1)>

**Cluster 2:**

12. petallength='(3.36-4.54]' 34 ==> cluster=cluster2 30 <conf:(0.88)>

**Cluster 3:**

1. petallength='(-inf-2.18]' 50 ==> cluster=cluster3 50 <conf:(1)>

In this case also because cluster have only 35 instances it can not support rules with more than 35/150 support

