Advance Data Mining-Association Analysis -1

Roshni Sundaramurthy(rossu809), Yusur Al-Mter (yusal621)

05 March 2019

Association Analysis - 1

Dataset

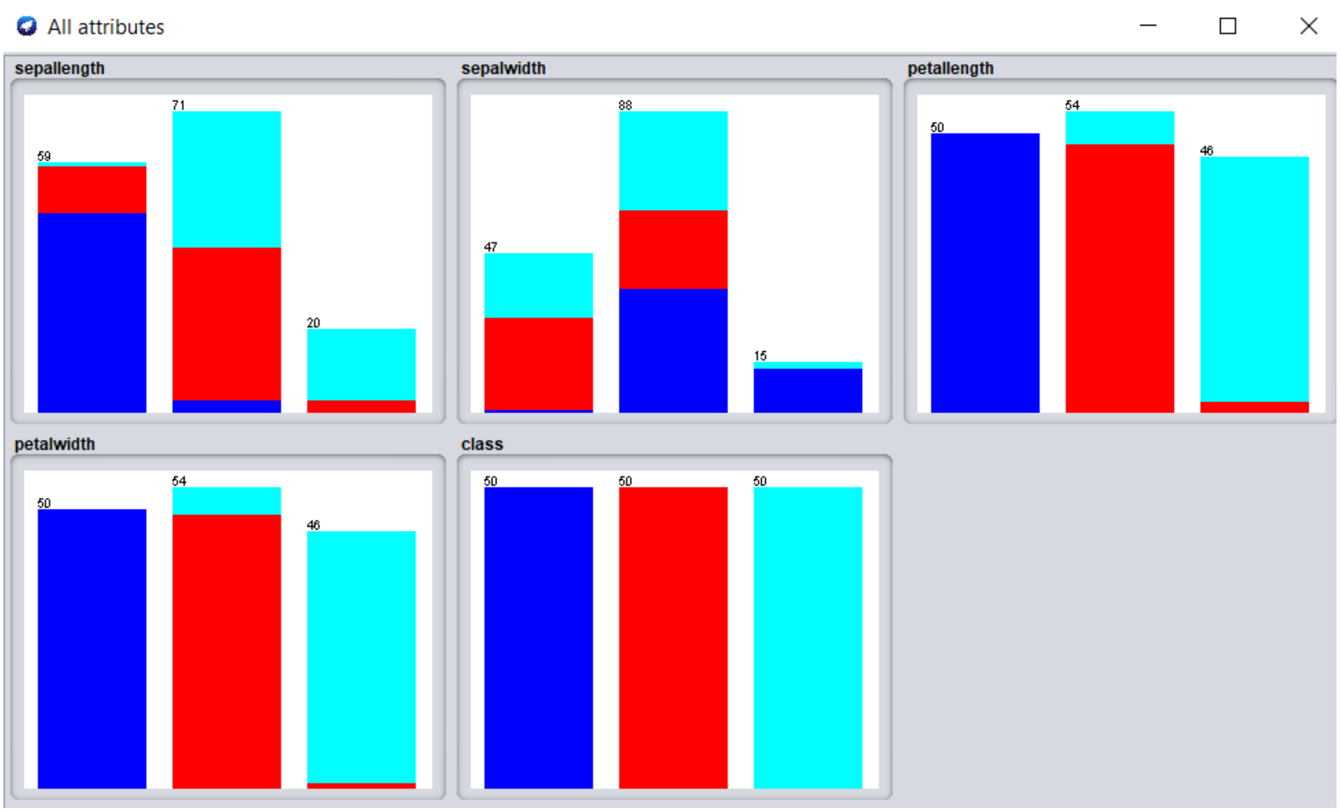
The dataset consists of 50 samples from each of three species of Iris flowers (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample, they are the length and the width of sepal and petal.

Goals

Cluster a given dataset and use association analysis to describe the clusters obtained.

Discretizing

Since the association analysis in Weka (Apriori algorithm) cannot cope with continuous attributes, we should discretize the iris dataset with 3 bins (number of states of the discretized attributes).



Clustering

Applied SimpleKmeans clusterer to the data with 3 clusters (since we know there are 3 types of Iris flowers) and seed value 10.

```
Attributes:   5
              sepallength
              sepalwidth
              petallength
              petalwidth
Ignored:
              class
Test mode:    Classes to clusters evaluation on training data

=== Clustering model (full training set) ===


kMeans
======

Number of iterations: 3
Within cluster sum of squared errors: 96.0

Initial starting points (random):

Cluster 0: '\'(5.5-6.7]\'','\'(2.8-3.6]\'','\'(2.966667-4.933333]\'','\'(0.9-1.7]\''
Cluster 1: '\'(6.7-inf)\'','\'(2.8-3.6]\'','\'(4.933333-inf)\'','\'(1.7-inf)\''
Cluster 2: '\'(-inf-5.5]\'','\'(3.6-inf)\'','\'(-inf-2.966667]\'','\'(-inf-0.9]\''

Missing values globally replaced with mean/mode

Final cluster centroids:
                                                        Cluster#
Attribute                             Full Data                0                1                2
                                        (150.0)           (55.0)           (45.0)           (50.0)
======================================================================================================
sepallength                       '(5.5-6.7]'      '(5.5-6.7]'      '(5.5-6.7]'      '(-inf-5.5]'
sepalwidth                        '(2.8-3.6]'      '(-inf-2.8]'     '(2.8-3.6]'      '(2.8-3.6]'
petallength              '(2.966667-4.933333]' '(2.966667-4.933333]' '(4.933333-inf)' '(-inf-2.966667]'
petalwidth                        '(0.9-1.7]'      '(0.9-1.7]'      '(1.7-inf)'      '(-inf-0.9]'
```

```
Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       55 ( 37%)
1       45 ( 30%)
2       50 ( 33%)


Class attribute: class
Classes to Clusters:

  0  1  2  <-- assigned to cluster
  0  0 50 | Iris-setosa
 48  2  0 | Iris-versicolor
  7 43  0 | Iris-virginica

Cluster 0 <-- Iris-versicolor
Cluster 1 <-- Iris-virginica
Cluster 2 <-- Iris-setosa

Incorrectly clustered instances :       9.0       6      %
```

Sum of squared error is 96. Incorrectly clustered instances seems to be 9 which indicates that the class Iris-Setosa is clusterd correctly. It is assigned to 2nd cluster with perfection when compared to other 2 pairs.

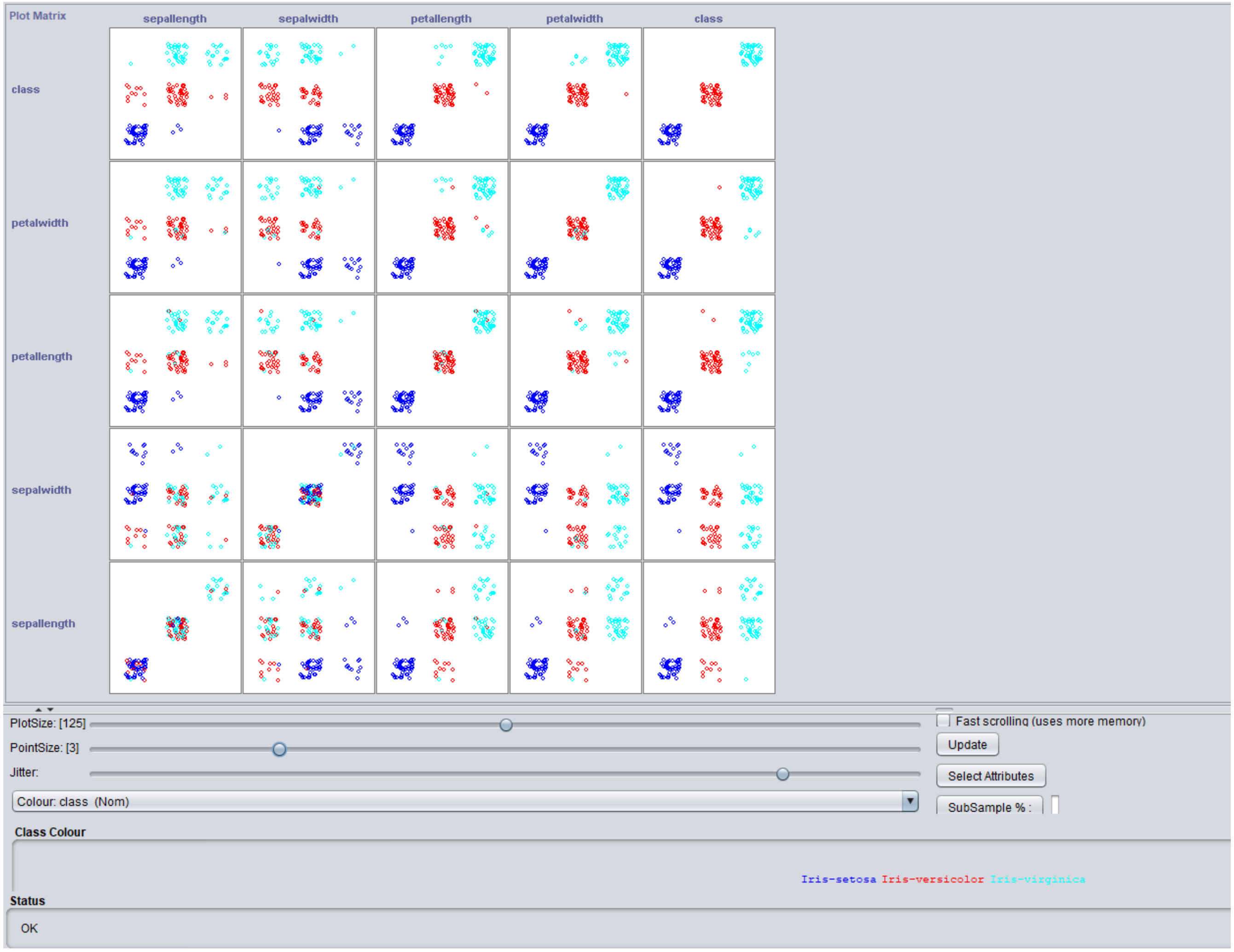
Association analysis

We used the apriori algorithm for association analysis with minimum metric=0.9 and number of rules=10. The association rules have numbers on the right and left hand sides of the conjunctions of attribute-value pairs of each rule and they indicates the support of the determinant and of the determinant plus the consequent.

```
Apriori
=======

Minimum support: 0.3 (45 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13

Size of set of large itemsets L(2): 10

Size of set of large itemsets L(3): 5

Size of set of large itemsets L(4): 1

Best rules found:

 1. petalwidth='(-inf-0.9]' 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 2. petallength='(-inf-2.966667]' 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 3. class=Iris-setosa 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 4. petallength='(-inf-2.966667]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 5. class=Iris-setosa 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 6. petalwidth='(-inf-0.9]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 7. petalwidth='(-inf-0.9]' class=Iris-setosa 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 8. petallength='(-inf-2.966667]' class=Iris-setosa 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 9. petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
10. class=Iris-setosa 50 ==> petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
```

Visualization

We can see the data crosstabulated for each pair of attributes.



Describing clustering through association analysis

A new attribute is created and it has been created with the clustering label. Again we repeated the above same steps. We ran the apriori algorithm with SimpleKmeans with number of clusters=3.

```
Instances:    150
Attributes:   6
              sepallength
              sepalwidth
              petallength
              petalwidth
              class
              cluster
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.3 (45 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 16

Size of set of large itemsets L(2): 17

Size of set of large itemsets L(3): 14

Size of set of large itemsets L(4): 6

Size of set of large itemsets L(5): 1

Best rules found:

 1. petalwidth='(-inf-0.9]' 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 2. petallength='(-inf-2.966667]' 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 3. class=Iris-setosa 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 4. petallength='(-inf-2.966667]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 5. cluster=cluster3 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 6. petallength='(-inf-2.966667]' 50 ==> cluster=cluster3 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 7. class=Iris-setosa 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 8. petalwidth='(-inf-0.9]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 9. cluster=cluster3 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
10. petalwidth='(-inf-0.9]' 50 ==> cluster=cluster3 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
```

Accurate rules for three clusters:

The rules that are accurate and such that the antecedent does not contain the class attribute and the consequent only contains the cluster attribute are displayed below:

Rule 6 and 10:

1. petallength='(-inf-2.966667]' 50 ==> cluster=cluster3 50
2. petalwidth='(-inf-0.9]' 50 ==> cluster=cluster3 50

3 possible variations:

Different clustering algorithm

Repeating the above steps with simple EM algorithm with clusters=3, seed value=10, min standard deviation = 1.0E-6

```
EM
==

Number of clusters: 3
Number of iterations performed: 0


                      Cluster
Attribute                  0    1    2
                        (0.37) (0.33) (0.3)
===================================
sepallength
  '(-inf-5.5]'            13   48    1
  '(5.5-6.7]'             41    4   29
  '(6.7-inf)'              4    1   18
  [total]                 58   53   48
sepalwidth
  '(-inf-2.8]'            35    2   13
  '(2.8-3.6]'             22   37   32
  '(3.6-inf)'              1   14    3
  [total]                 58   53   48
petallength
  '(-inf-2.966667]'        1   51    1
  '(2.966667-4.933333]'   52    1    4
  '(4.933333-inf)'         5    1   43
  [total]                 58   53   48
petalwidth
  '(-inf-0.9]'             1   51    1
  '(0.9-1.7]'             53    1    3
  '(1.7-inf)'              4    1   44
  [total]                 58   53   48
```

```
Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       56 ( 37%)
1       50 ( 33%)
2       44 ( 29%)


Log likelihood: -2.67073


Class attribute: class
Classes to Clusters:

  0  1  2  <-- assigned to cluster
  0 50  0 | Iris-setosa
 49  0  1 | Iris-versicolor
  7  0 43 | Iris-virginica

Cluster 0 <-- Iris-versicolor
Cluster 1 <-- Iris-setosa
Cluster 2 <-- Iris-virginica

Incorrectly clustered instances :       8.0       5.3333 %
```

Incorrectly clustered instances seems to be 8. Iris-Setosa class is assigned to 1st cluster. But one instance of

Iris-Versicolor class is incorrectly clustered and assigned to 2nd cluster. It seems better than SimpleKmeans clusterer.

Association analysis with 5 attributes

```
Apriori
=======

Minimum support: 0.3 (45 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13

Size of set of large itemsets L(2): 10

Size of set of large itemsets L(3): 5

Size of set of large itemsets L(4): 1

Best rules found:

 1. petalwidth='(-inf-0.9]' 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 2. petallength='(-inf-2.966667]' 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 3. class=Iris-setosa 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 4. petallength='(-inf-2.966667]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 5. class=Iris-setosa 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 6. petalwidth='(-inf-0.9]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 7. petalwidth='(-inf-0.9]' class=Iris-setosa 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 8. petallength='(-inf-2.966667]' class=Iris-setosa 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 9. petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
10. class=Iris-setosa 50 ==> petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
```

Association analysis with 6 attributes (along with cluster attribute)

```
Apriori
=======

Minimum support: 0.3 (45 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 15

Size of set of large itemsets L(2): 17

Size of set of large itemsets L(3): 14

Size of set of large itemsets L(4): 6

Size of set of large itemsets L(5): 1

Best rules found:

 1. petalwidth='(-inf-0.9]' 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 2. petallength='(-inf-2.966667]' 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 3. class=Iris-setosa 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 4. petallength='(-inf-2.966667]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 5. cluster=cluster2 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 6. petallength='(-inf-2.966667]' 50 ==> cluster=cluster2 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 7. class=Iris-setosa 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 8. petalwidth='(-inf-0.9]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 9. cluster=cluster2 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
10. petalwidth='(-inf-0.9]' 50 ==> cluster=cluster2 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
```

Minimum support=0.3 (45 instances)

Accurate rules:

Rule 6 and 10:

1. petallength='(-inf-2.966667]' 50 ==> cluster=cluster2 50
2. petalwidth='(-inf-0.9]' 50 ==> cluster=cluster2 50

Different number of clusters

Now, number of clusters is choosen as 5 and applying EM algorithm with seed value=10, number of bins=3.

Clustering

```
Attributes:    6
               sepallength
               sepalwidth
               petallength
               petalwidth
Ignored:
               class
               cluster
Test mode:     Classes to clusters evaluation on training data

=== Clustering model (full training set) ===


EM
==

Number of clusters: 5
Number of iterations performed: 0


                      Cluster
Attribute                  0   1   2   3   4
                       (0.23) (0.17) (0.33) (0.09) (0.17)
===========================================
sepallength
  '(-inf-5.5]'            12   2  48   1   1
  '(5.5-6.7]'             23  23   4  10  16
  '(6.7-inf)'              2   4   1   6  12
  [total]                 37  29  53  17  29
sepalwidth
  '(-inf-2.8]'            35   1   2  13   1
  '(2.8-3.6]'              1  27  37   1  27
  '(3.6-inf)'              1   1  14   3   1
  [total]                 37  29  53  17  29
petallength
  '(-inf-2.966667]'        1   1  51   1   1
  '(2.966667-4.933333]'   31  25   1   1   1
  '(4.933333-inf)'         5   3   1  15  27
  [total]                 37  29  53  17  29
petalwidth
  '(-inf-0.9]'             1   1  51   1   1
  '(0.9-1.7]'             32  24   1   1   1
  '(1.7-inf)'              4   4   1  15  27
  [total]                 37  29  53  17  29
```

```
Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       34 ( 23%)
1       25 ( 17%)
2       50 ( 33%)
3       14 (  9%)
4       27 ( 18%)


Log likelihood: -2.68013


Class attribute: cluster
Classes to Clusters:

  0  1  2  3  4  <-- assigned to cluster
 34  0  0  0  0 | cluster1
  0 25  0  0  0 | cluster2
  0  0 50  0  0 | cluster3
  0  0  0 14  0 | cluster4
  0  0  0  0 27 | cluster5

Cluster 0 <-- cluster1
Cluster 1 <-- cluster2
Cluster 2 <-- cluster3
Cluster 3 <-- cluster4
Cluster 4 <-- cluster5

Incorrectly clustered instances :       0.0       0      %
```

We can see that the instances are clustered well in every clusters.

Association analysis with 5 attributes

```
=== Run information ===

Scheme:       weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:     iris-weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R1-4-precision6
Instances:    150
Attributes:   5
              sepallength
              sepalwidth
              petallength
              petalwidth
              class
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.3 (45 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13

Size of set of large itemsets L(2): 10

Size of set of large itemsets L(3): 5

Size of set of large itemsets L(4): 1

Best rules found:

 1. petalwidth='(-inf-0.9]' 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 2. petallength='(-inf-2.966667]' 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 3. class=Iris-setosa 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 4. petallength='(-inf-2.966667]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 5. class=Iris-setosa 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 6. petalwidth='(-inf-0.9]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 7. petalwidth='(-inf-0.9]' class=Iris-setosa 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 8. petallength='(-inf-2.966667]' class=Iris-setosa 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 9. petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
10. class=Iris-setosa 50 ==> petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
```

Association analysis with 6 attributes (along with cluster attribute)

```
Attributes:   6
              sepallength
              sepalwidth
              petallength
              petalwidth
              class
              cluster
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.3 (45 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 14

Size of set of large itemsets L(2): 14

Size of set of large itemsets L(3): 11

Size of set of large itemsets L(4): 5

Size of set of large itemsets L(5): 1

Best rules found:

 1. petalwidth='(-inf-0.9]' 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 2. petallength='(-inf-2.966667]' 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 3. class=Iris-setosa 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 4. petallength='(-inf-2.966667]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 5. cluster=cluster5 50 ==> petallength='(-inf-2.966667]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 6. petallength='(-inf-2.966667]' 50 ==> cluster=cluster5 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 7. class=Iris-setosa 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 8. petalwidth='(-inf-0.9]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 9. cluster=cluster5 50 ==> petalwidth='(-inf-0.9]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
10. petalwidth='(-inf-0.9]' 50 ==> cluster=cluster5 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
```

Accurate rules:

Rule 6 and 10:

1. petallength='(-inf-2.966667]' 50 ==> cluster=cluster5 50
2. petalwidth='(-inf-0.9]' 50 ==> cluster=cluster5 50

Different number of bins

Number of clusters = 5, number of bins = 5 and applying Hierarchical clustering algorithm with seed value=10.

Clustering

```
Time taken to build model (full training data) : 0.03 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       50 ( 33%)
1        6 (  4%)
2       90 ( 60%)
3        3 (  2%)
4        1 (  1%)


Class attribute: cluster
Classes to Clusters:

  0  1  2  3  4  <-- assigned to cluster
 50  0  0  0  0 | cluster1
  0  6  0  0  0 | cluster2
  0  0 90  0  0 | cluster3
  0  0  0  3  0 | cluster4
  0  0  0  0  1 | cluster5

Cluster 0 <-- cluster1
Cluster 1 <-- cluster2
Cluster 2 <-- cluster3
Cluster 3 <-- cluster4
Cluster 4 <-- cluster5

Incorrectly clustered instances :       0.0       0      %
```

We can see that the instances are clustered such that more instances fall in 3rd cluster, only 3 in 4th cluster and only 1 in 5th cluster.

Association analysis with 5 attributes

```
Apriori
=======

Minimum support: 0.3 (45 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 8

Size of set of large itemsets L(2): 3

Size of set of large itemsets L(3): 1

Best rules found:

 1. class=Iris-setosa 50 ==> petallength='(-inf-2.18]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 2. petallength='(-inf-2.18]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 3. petalwidth='(-inf-0.58]' 49 ==> petallength='(-inf-2.18]' 49    <conf:(1)> lift:(3) lev:(0.22) [32] conv:(32.67)
 4. petalwidth='(-inf-0.58]' 49 ==> class=Iris-setosa 49    <conf:(1)> lift:(3) lev:(0.22) [32] conv:(32.67)
 5. petalwidth='(-inf-0.58]' class=Iris-setosa 49 ==> petallength='(-inf-2.18]' 49    <conf:(1)> lift:(3) lev:(0.22) [32] conv:(32.67)
 6. petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' 49 ==> class=Iris-setosa 49    <conf:(1)> lift:(3) lev:(0.22) [32] conv:(32.67)
 7. petalwidth='(-inf-0.58]' 49 ==> petallength='(-inf-2.18]' class=Iris-setosa 49    <conf:(1)> lift:(3) lev:(0.22) [32] conv:(32.67)
 8. petallength='(-inf-2.18]' 50 ==> petalwidth='(-inf-0.58]' 49    <conf:(0.98)> lift:(3) lev:(0.22) [32] conv:(16.83)
 9. class=Iris-setosa 50 ==> petalwidth='(-inf-0.58]' 49    <conf:(0.98)> lift:(3) lev:(0.22) [32] conv:(16.83)
10. petallength='(-inf-2.18]' class=Iris-setosa 50 ==> petalwidth='(-inf-0.58]' 49    <conf:(0.98)> lift:(3) lev:(0.22) [32] conv:(16.83)
```

Association analysis with 6 attributes (along with cluster attribute)

```
Attributes:   6
              sepallength
              sepalwidth
              petallength
              petalwidth
              class
              cluster
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.3 (45 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 10

Size of set of large itemsets L(2): 7

Size of set of large itemsets L(3): 4

Size of set of large itemsets L(4): 1

Best rules found:

 1. class=Iris-setosa 50 ==> petallength='(-inf-2.18]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 2. petallength='(-inf-2.18]' 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 3. cluster=cluster1 50 ==> petallength='(-inf-2.18]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 4. petallength='(-inf-2.18]' 50 ==> cluster=cluster1 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 5. cluster=cluster1 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 6. class=Iris-setosa 50 ==> cluster=cluster1 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 7. class=Iris-setosa cluster=cluster1 50 ==> petallength='(-inf-2.18]' 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 8. petallength='(-inf-2.18]' cluster=cluster1 50 ==> class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
 9. petallength='(-inf-2.18]' class=Iris-setosa 50 ==> cluster=cluster1 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
10. cluster=cluster1 50 ==> petallength='(-inf-2.18]' class=Iris-setosa 50    <conf:(1)> lift:(3) lev:(0.22) [33] conv:(33.33)
```

Accurate rules:

Rule 4 and 9:

1. petallength='(-inf-2.18]' 50 ==> cluster=cluster1 50 and this belogs to class Iris-Setosa.

Conclusion

So far we tried possible variations and found some results. It seems that the instances are clusterd well when the number of clusters is 3. Also EM with 3 clusters seems to be better than Kmeans with 3 clusters in terms of accuracy. The rules are choosen such that it has 100% confidence with minimum support. When we used 5 clusters, very less instances tend to fall in several clusters. For example, one instance in one cluster. So, based on the data, we need to choose enough number of clusters.