

Name: Dhiraj Ravindra Bodake
Roll No: 18141216

Title : Experiment with clustering simple k Means
on weka.

Theory :

Simple k-means clustering - while this dataset is commonly used to test classification algorithms, we will experiment here to see how well the k-means clustering algorithm clusters the numeric data according to the original class labels.

click the "cluster" tab at the top of the Weka Explorer.

click the clusterer "choose" button and select "simple k means".

click the Simple k means command box to the right of the choose button, change the "numclusters" attribute to 3, and click the ok button. Under cluster mode, select the radio button "classes to cluster evaluation" which should be follow by "(nom) class" by default. Press start to begin k-means clustering and evaluation.

- Note the cluster centroids in the clusterer output pane

- what instance percentage is incorrectly clustered?

- visualization In k-means clustering, there are a number of ways one can often improve results.

one of the most common is to normalize the results in some fashion so that the difference in scale of the numerical attributes do not dominate the Euclidean distance measure

This can be accomplished by linearly scaling the data of each attribute between -1 and 1, or by replacing attribute values with the number of standard deviations each have from the attribute mean value. For this dataset, we will demonstrate an even simpler approach.

Visualization can sometimes help us discern the attributes that best separate the data. Recall that k-means clustering assumes nonoverlapping, hyperspherical clusters with similar size and density.

Each subplot shows the normalized value distribution of a pair of attributes, colored by class.

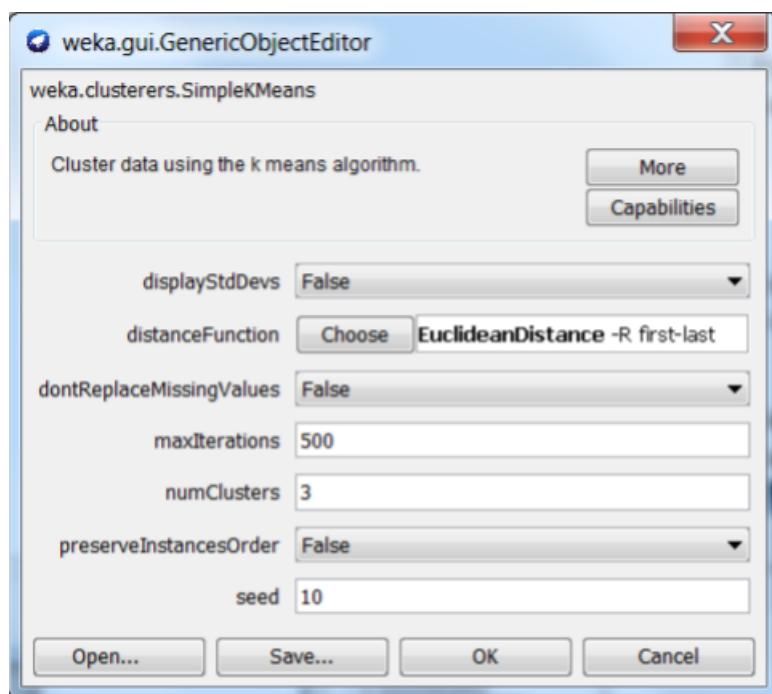
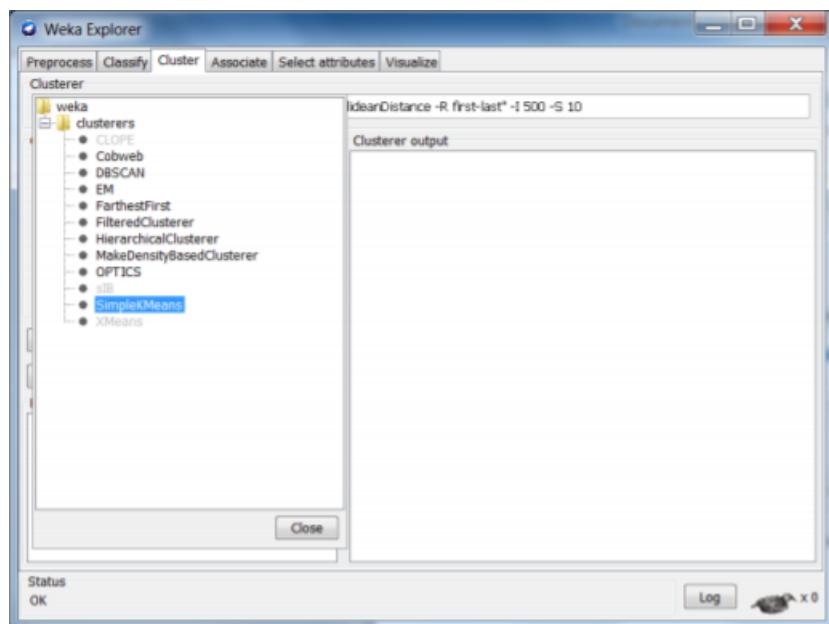
Adjust the visualization settings: plotsize increases the size of each subplot.

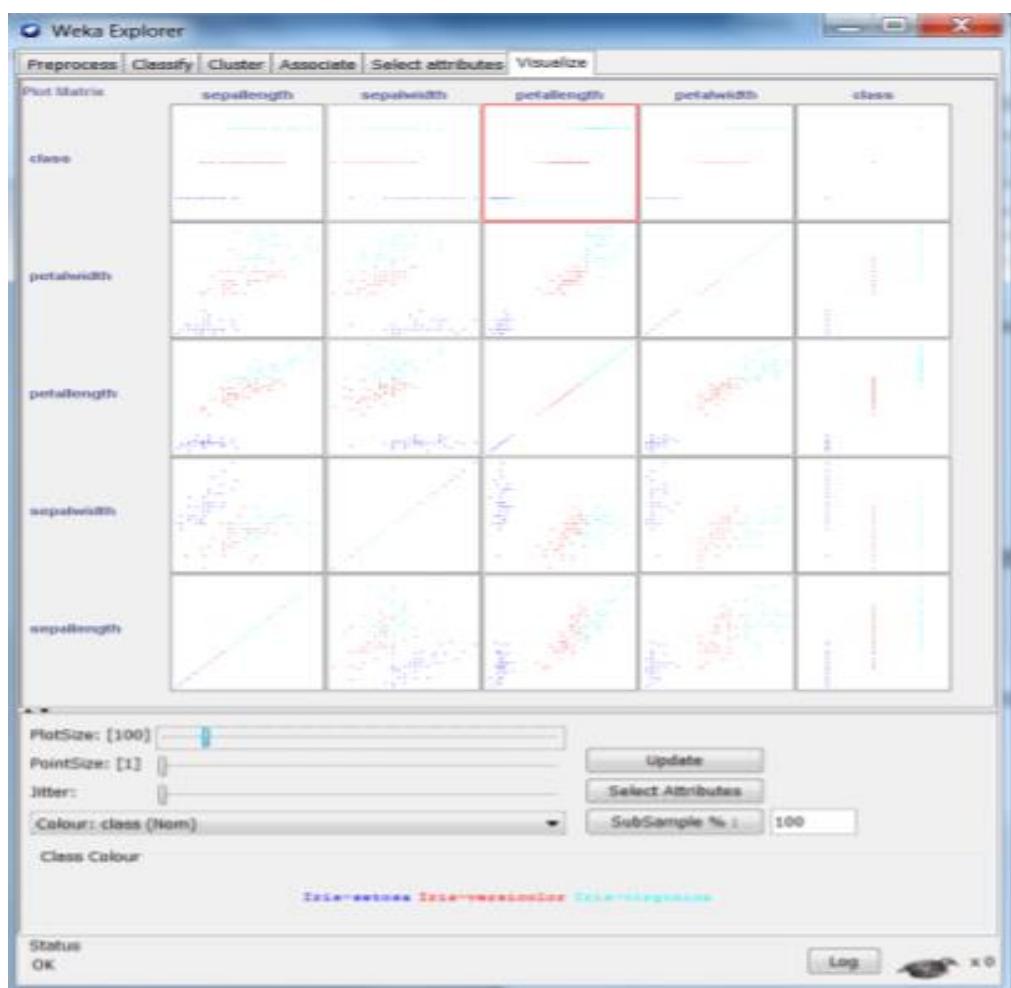
- o point size makes the individual point large.
- o jitter randomly shifts the point so the overlapping point can be seen and density become more visually apparent.

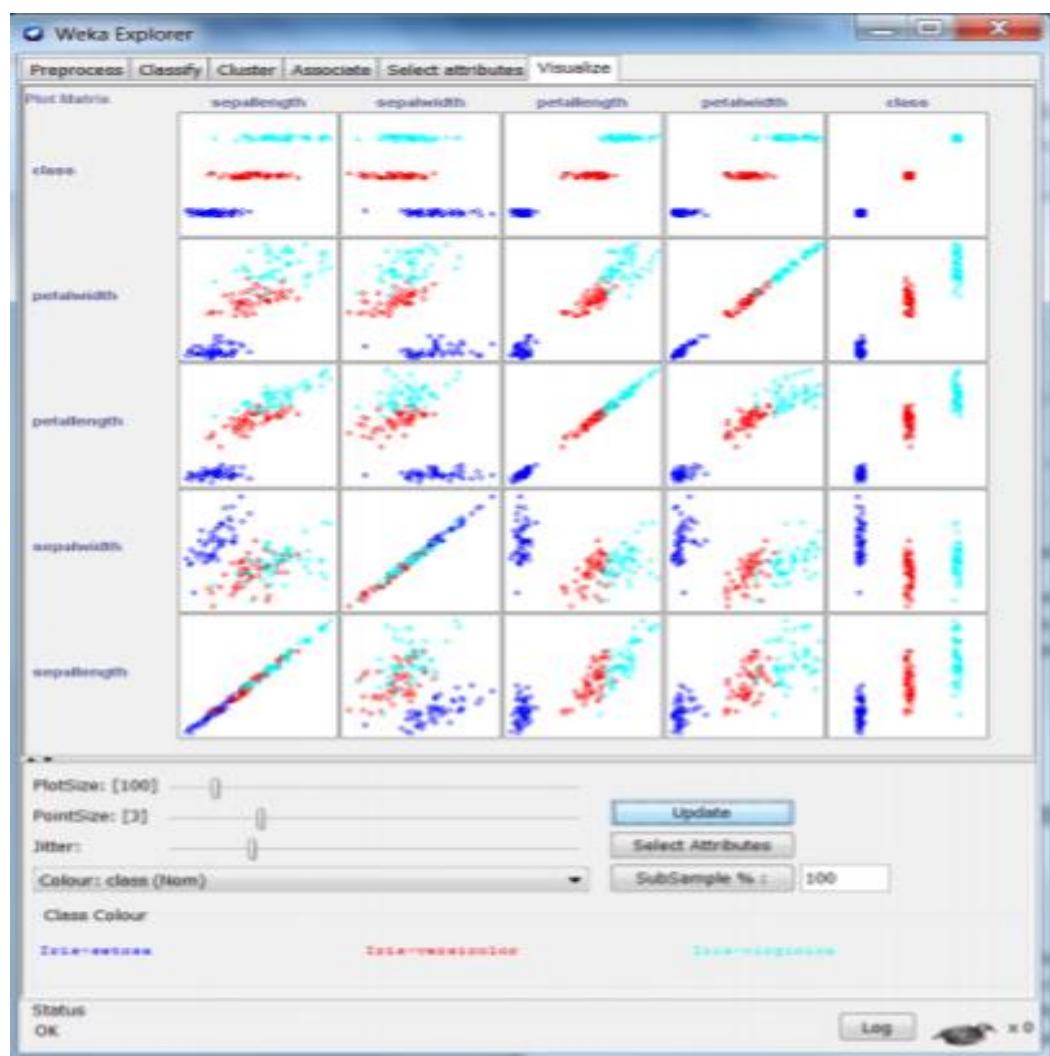
Simple K-means clustering Ignoring Attributes

- click on the toggle the attributes dark
- ignore the two numeric attributes that don't separates as well as the ones you have written above. Press start to begin k-means clustering and evaluation.

Because k-means clustering assumes non-overlapping, hyperspherical clusters of data with similar size and density, data attributes that violate this assumption can be detrimental to clustering performance. less is sometimes more. This is why the use of visualization tools can be helpful in the best application of clustering algorithms







Conclusion:

Thus We studied and understood the Clustering Simple K Means on weka.