**CURE Algorithm**

**Проект по „Извличане на закономерности от данни”**

**на Георги Тончев фн:** **М23449 , Светлин Чобанов фн: М23459**

**и Диана Чалъкова фн: М23337**

**Cure:**

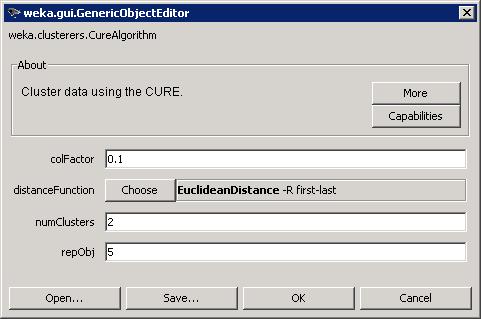
The CURE method is an example of agglomerative clustering. The main difference with other such algorithms is the "smooth" selection of representative points at any rate. As a result of the selection, the "irregular" form of representational points may be broken.

Method of use:

The algorithm is integrated into the already existing Weka system and uses its own interface. The system provides the ability to select a large part of the parameters in your interface.

Required parameters:

1. repObj - maximum number of representative objects in a cluster
2. numClusters - desire number of output clusters
3. colFactor- Bending factor (in the range of 0 to 1)
4. distanceFunction- function of spacing between points in a cluster.



Description of the algorithm**:**

1. Initialize the cluster set (all point - separate cluster).

2. For all clusters Ci intentions nearest-cluster Cj (at this stage of dissolution between clusters is simply the distance between the corresponding points of the thesis clusters)

3. Plug in the two closest clusters Ck and Cl in cluster Cm (as ever Cm is a unit of points from clusters Ck and Cl)

4For cluster Cm select c representative (well distributed in space) points:

i. Select the first point, farthest from the center of Cm

ii. As long as the d points are not selected devices:

Select the point that is farthest from the previous one

5. Move the selected c points perpendicular to the center of the cluster s

bending factor α: p = p + α (mean - p), where p is the coordinates of

the representation point, mean - is the coordinates of the centroid.

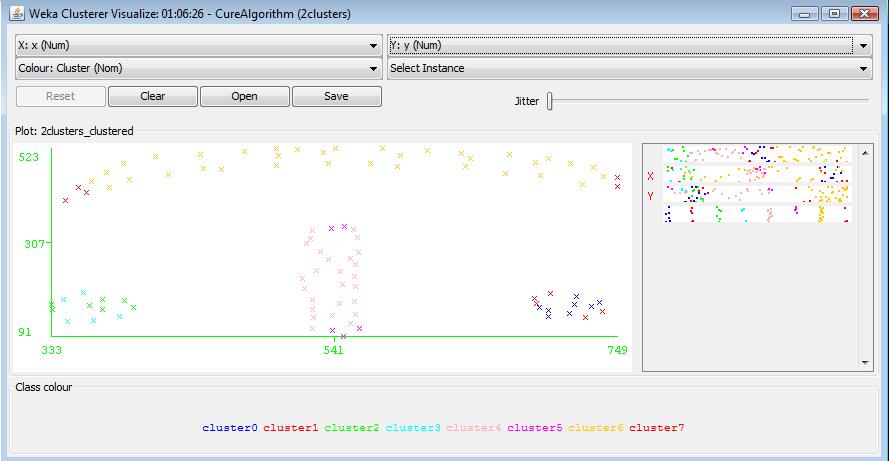
6. Determine the cluster, which is closest to Cm, as the distance between clusters are determined by the distance between two closest representations points on the corresponding clusters.

7. Plush the nearest pair of clusters

8. Repeat steps 3-7 until no cluster remains.



Figure (a) represents two clusters as the larger crosses represent the representation points. In this iteration, the algorithm will merge the two clusters as the closest ones (figure (b)), because the distance between the two representative points of the line is the smallest.



Divide one hundred points into four clusters with CURE. One by one, starting from cluster number 1, the representative points of the respective clusters are depicted. It can be seen that the dots do not form a proper geometric figure.

Peculiarities of implementation**:**

1. 1When selecting representative points from a cluster, if there are more than a number of points with equal spacing - theses are selected that entered the cluster earlier (at the earliest possible iteration).
2. 2) The algorithm works with multi-dimensional data.