Data Mining Homework 6

Madis Nõmme

March 27th, 2014

Task 1

Implemented k-means clustering algorithm in Ruby. The resulting clusters:

```
Cluster:   (6.6000000000000005,3.8000000000000003)
  Members: [(7,3), (5,3), (7,4), (6,5), (8,4)]
Cluster:   (2.5,5.25)
  Members: [(2,4), (3,5), (2,5), (3,7)]
Cluster:   (6.0,8.0)
  Members: [(6,8)]
```

Same thing for k-medoids:

Task 2

Got 3 new clusterings that are different from the original:

1. Starting: [, ,] Cluster: (8.45,9.05) Members: [(6,8)] Cluster: (-0.7,4.6) Members: [(2,4), (3,5), (2,5), (3,7)] Cluster: (4.8,0.1) Members: [(7,3), (5,3), (7,4), (6,5), (8,4)]
2. Starting: [, ,] Cluster: (6.2,2.4166666666666665) Members: [(7,3), (5,3), (7,4), (6,5), (8,4)] Cluster: (-5.0,5.6) Members: [(2,4), (3,5), (6,8), (2,5), (3,7)] Cluster: (3.7,-5.8) Members: []
3. Starting: [, ,] Cluster: (4.8,-9.3) Members: [] Cluster: (4.945454545454545,3.6636363636363636) Members: [(2,4), (7,3), (3,5), (5,3), (7,4), (6,8), (6,5), (8,4), (2,5), (3,7)] Cluster: (-7.0,-5.8) Members: []

Task 3

It happens because some points of my imagined 'real clustering' are closer to the other cluster center than to their rightful cluster center. This is by definition of the algorithm and correct.

Task 4

To remedy the situation in Task 3, I would suggest having more than 1 *center* point of a cluster. The criteria for a point belonging to a cluster would be that it's close-enough to any of the cluster *centers*.

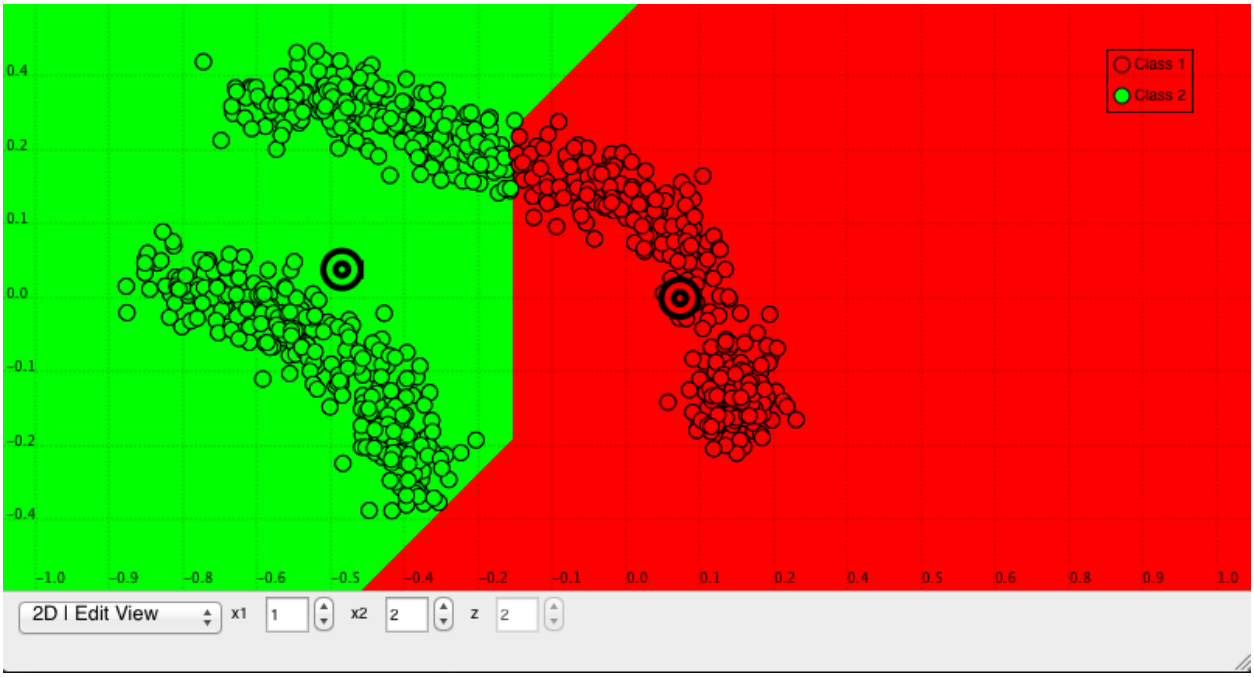

Figure 1: k-means not working