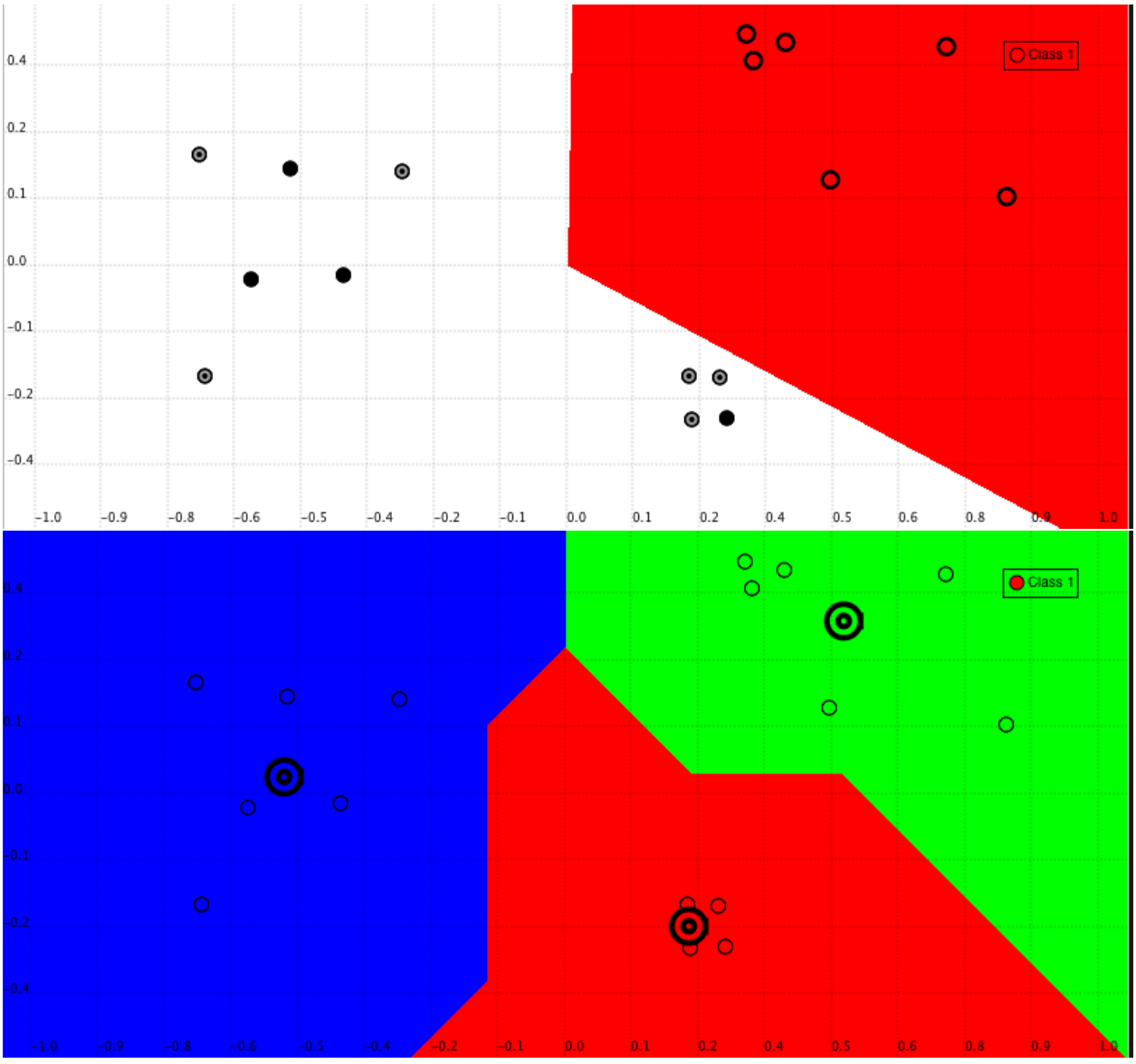
Data Mining Homework 7

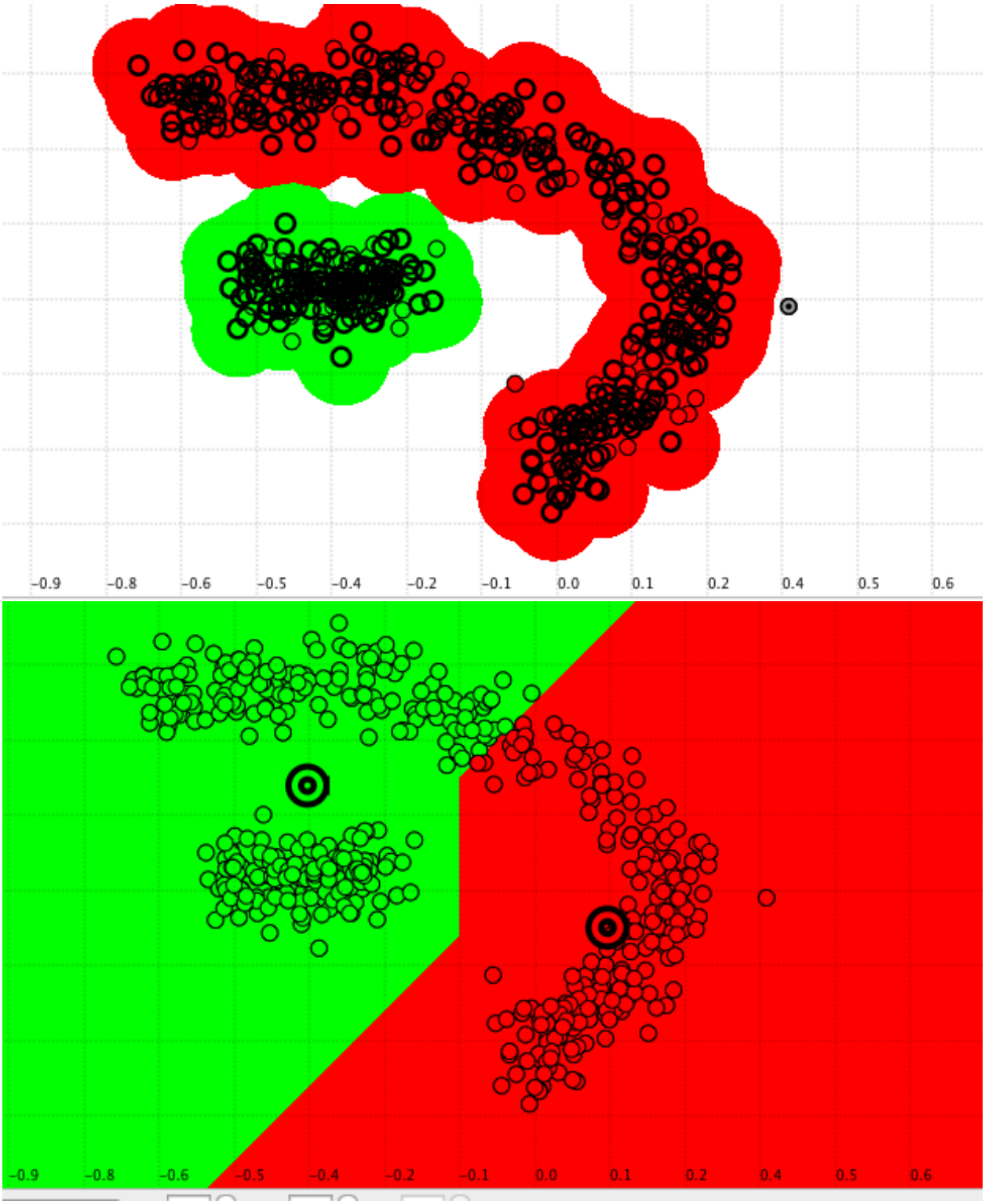
Madis Nõmme

April 3rd, 2014

1 DBSCAN vs K-means

1.1 K-means





1.1.1 Pros

1. Easy to understand and implement
2. Fast

1.1.2 Cons

1. Can't find arbitrary shaped clusters
2. One has to know the number of clusters
3. Choosing initial cluster centers affects results

1.1.3 DBSCAN

1.1.4 Pros

1. No need to know cluster centers

1.1.5 Cons

1. Can't find clusters with highly varying density
2. Requires

2 Applying DBSCAN

Implemented algorithm in Ruby. It's a bit buggy. Questions for practice session:

- should the clusters be re-used or always created new ones?
- should points be re-considered for other clusters after already assigning them to some cluster?

1. Noise:
2. Border:
3. Core points:

3 Self organizing maps for pixel data

One-way seriation: only swap rows Two-way seriation: swap rows & columns

Do one-way seriation: define a goodness function, calculate current goodness (for the rows), re-order using some method (sorting, random, cluster), re-calulate goodness, if went better, re-apply.

shuffled.txt contains array of triplets. Triplet contains 3 values 0-255 of rgb.

```
rev0
 (0 0 0) (0 0 0) (0 0 0) etc.
```

4 Comparing Self-organizing maps with k-means and DBSCAN