Machine Learning Assignment 4

Mei Ling Wong 10-117-745

27 November 2013

Question 2

Run the k-means algorithm several times on the first data set with different initialisation. Do you always get the same solution? If not, why?

Answer
The result is depending on the initialized clusters. So running the algorithm with different initalization of the cluster centroids doesn't always give me the same solution, because the k-means algorithm achieves a local minimun and not a global one.

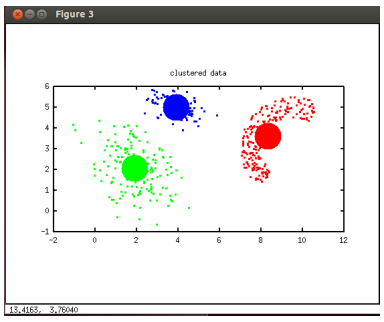

Figure 1: the first run of the algorithm

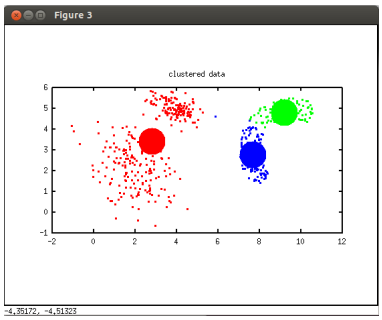

Figure 2: the second run of the algorithm

Question 3

You can see, that the k-means algorithm does not cluster the second dataset according to the ground truth. Why does this happen? Hint: The k-means algorithm is an optimisation method, it minimises a cost function. Calculate the cost function for the ground truth, and check, if it is optimal or not.

Answer
By the positions of the feature points, an half ring of points aroung a collection of points is formed. Just looking at this figure, it is clear that no matter how the cluster centroids can't be calculated so that the cost function is minimal.

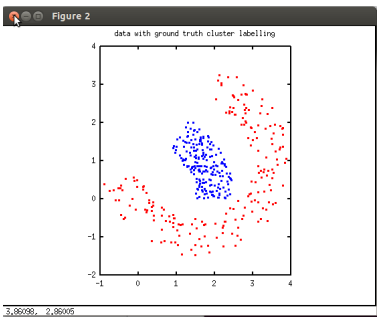

Figure 3: the ground truth

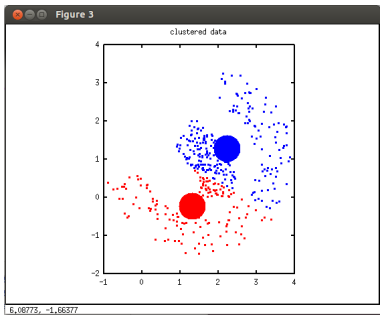

Figure 4: the result of the algorithm