Homework 4 Question 1

Yanlin Li, Student #: 1003770305

April 5, 2021

(a)

Here is the scatterplot:

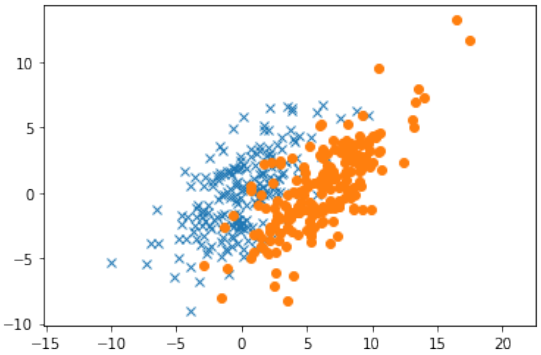

Figure 1: Scatterplot of the data

(b)

Here are the scatterplot and the cost vs. number of iteration plot:

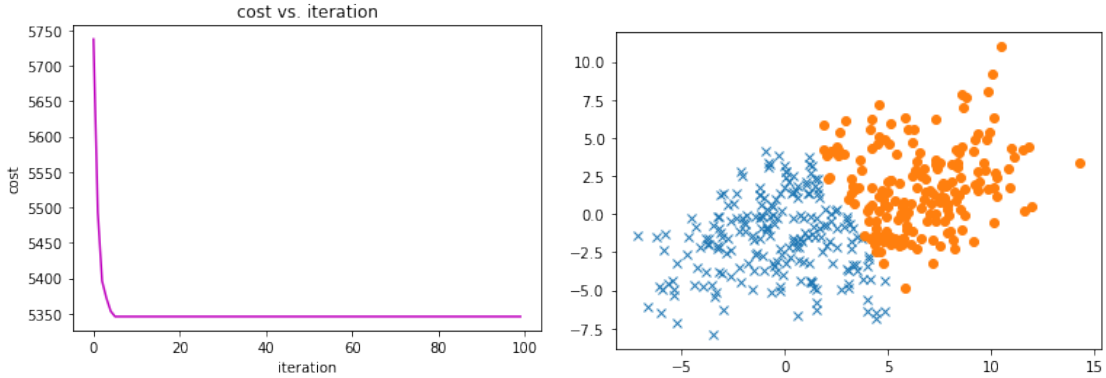

Figure 2: (a). cost vs. number of iterations, (b). Scatterplot of the data

The missclassification error is: 0.2625

(c)

Here are the scatterplot and the cost vs. number of iteration plot:

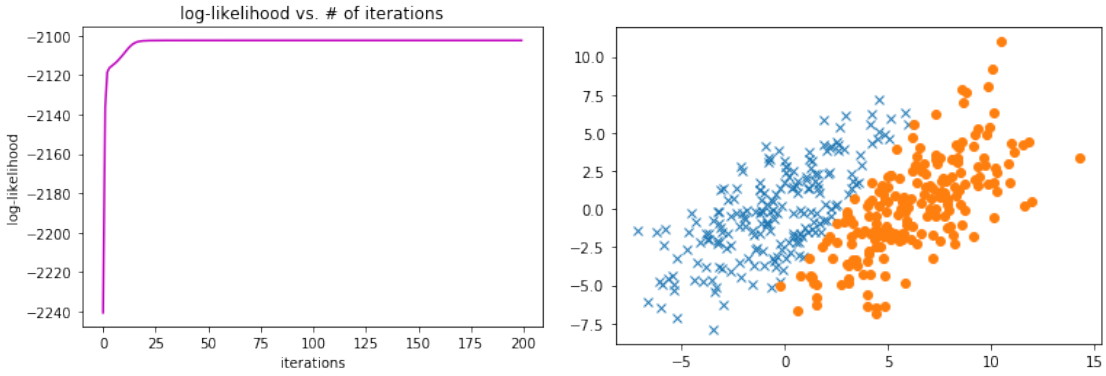

Figure 3: (a). log-likelihood vs. number of iterations, (b). Scatterplot of the data

The missclassification error is: 0.0825

(c)

1. EM algorithm gives a much more accurate classification. The missclassification error is lower and the scatterplot is more similar to the true one.

2. K-means converge faster than EM algorithm. The bottleneck of the two algorithms are about 5 for k-means and about 25 for EM algorithm.

3. The misclassification error of k-means are always higher than that of EM algorithm. So the performance does not depend on different realization of data.

The codes begin on the next page.