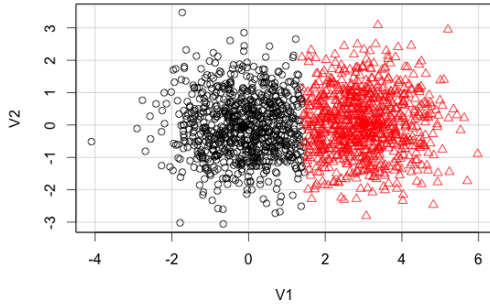


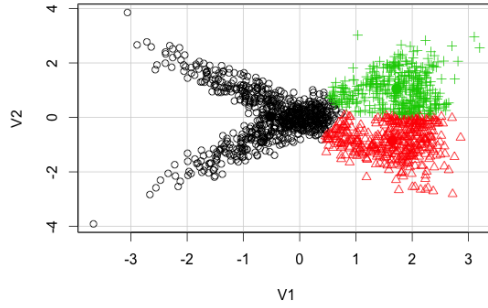
# Homework 3

Le Zhou

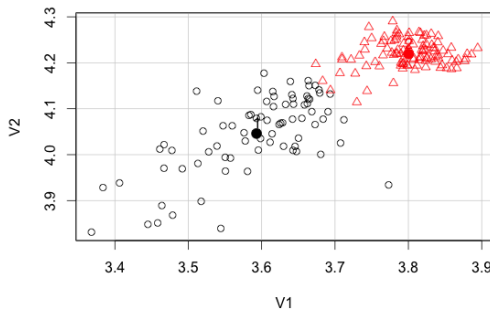
1:



dataset1



dataset2



dataset3

centers for dataset 1:

- (1) 3.0169857 0.04558442
- (2) -0.1308476 -0.05305959

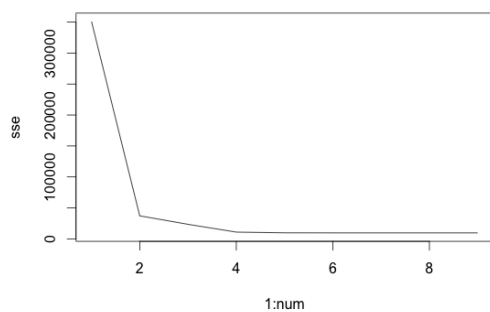
centers for dataset 2:

- (1) 1.6465040 -0.85083626
- (2) -0.5201668 0.03882836
- (3) 1.7247630 1.04927424

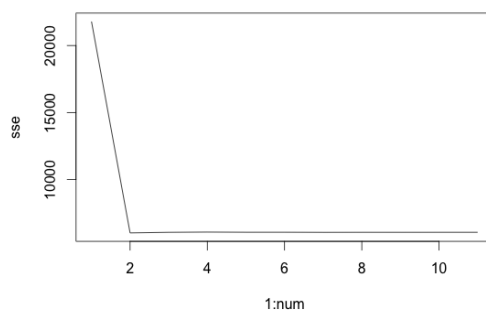
centers for dataset 3:

- (1) 3.800427 4.219084
- (2) 3.593044 4.046011

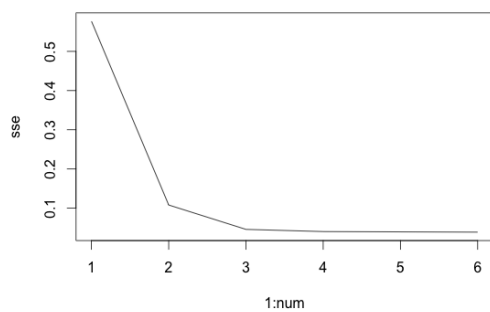
2:



dataset1

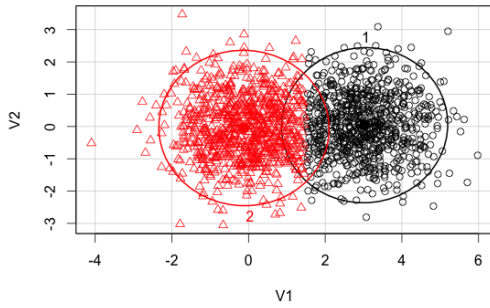


dataset2

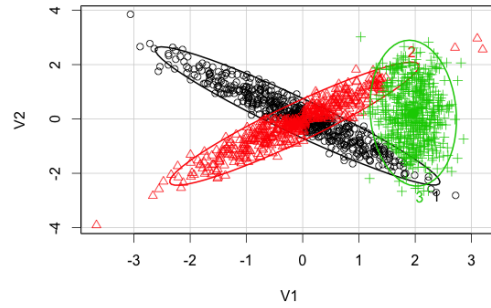


dataset3

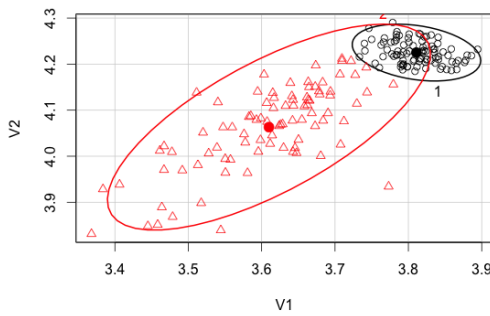
3:



dataset1



dataset2



dataset3

dataset 1:

k-means initialization

(1) 3.0269400 0.04039995

(2) -0.1211279 -0.04725860

EM

(1) 3.0289096 0.04007668

(2) -0.1191625 -0.04682670

dataset 2:

random initialization

(1) -0.182294 0.205905

(2) -1.395063 -1.151492

(3) 0.345710 0.507978

EM

(1) -0.0915112 0.1100919

(2) -0.1472908 -0.1744416

(3) 1.9645204 0.2128627

dataset3:

k-means initialization

(1) 3.800427 4.219084

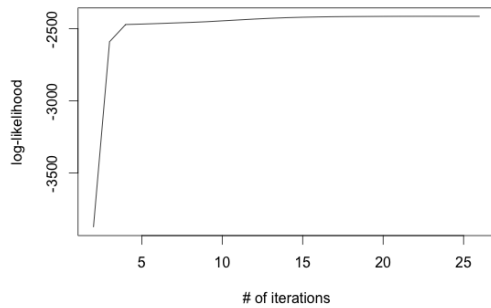
(2) 3.593044 4.046011

EM

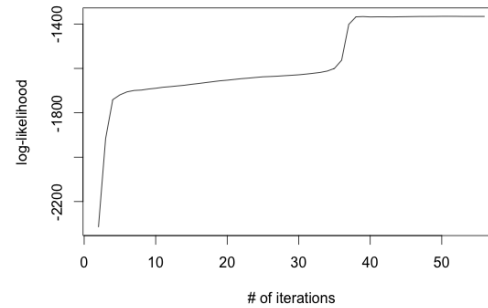
(1) 3.811004 4.224991

(2) 3.609953 4.063174

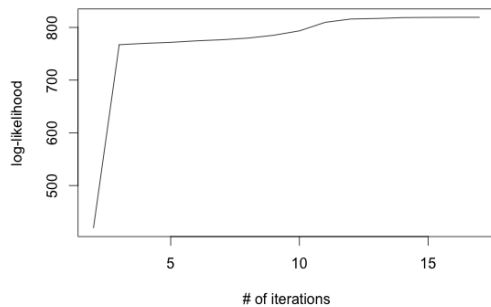
4:



dataset1



dataset2



dataset3

5:

For dataset 1 and 3, k-means and EM gives nearly the same results. This is because the clusters have only little overlap and the shape of clusters can be separated by a line. EM perform better than k-means on the second dataset for EM can separated complex clusters with irregular shape. However, EM has to try several initial values to get the best result.

**6:**

dataset1:

	1	2	3	4
bic	-2538.972	-2458.253	-2481.544	-2500.887
loglikelihhod	-2518.795	-2413.863	-2412.941	-2408.071

dataset2:

	1	2	3	4	5	6
bic	-1939.973	-1806.236	-1876.928	-1835.745	-1459.027	-1378.450
loglikelihhod	-1919.957	1762.201	-1808.874	-1743.671	-1342.935	-1238.339

dataset3:

	1	2	3	4	5
bic	722.5365	786.7543	773.6566	756.8992	745.3517
loglikelihood	737.2794	819.1887	823.7824	824.7165	830.8604

Log-likelihood should increase monotonous as k increase if the global optimum likelihood is achieved. However, EM depends on its initial points and can only find the local optimum. In some cases, like in dataset2, log-likelihood is not monotonous. In general, log-likelihood increases monotonous as k increase and BIC criteria perform well, especially on dataset 1 and 3.