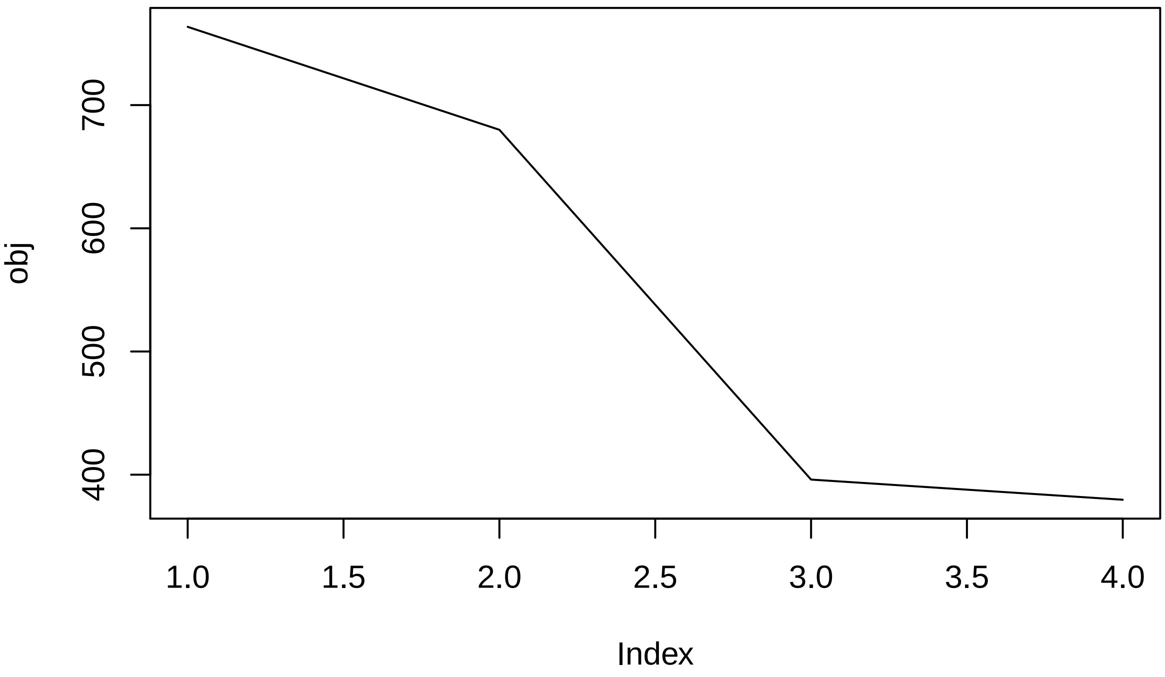
Homework 10

Jing Leng (GSI: Jiahe)

December 12, 2014

2

```
library(scatterplot3d)
mix <- read.table("mixgauss.dat")

c1 <- c(-2, -2, -4) # observed centers
c2 <- c(-4, -2, 2)
c3 <- c(4, 6, 2)
centroids <- rbind(c1, c2, c3)

classify <- function(x) {
  dis <- apply(centroids, 1, function(y) sqrt(sum((x - y)^2)));
  which(dis == min(dis))
}

label <- apply(mix, 1, classify )

label_old <- rep(1, 300)
obj = c()
while (1) {
  label <- apply(mix, 1, classify)
  if (sum(label != label_old) == 0) break
  centroids[1,] <- apply(mix[label==1,], 2, mean)
  centroids[2,] <- apply(mix[label==2,], 2, mean)
  centroids[3,] <- apply(mix[label==3,], 2, mean)
  label_old <- label
  obj = c(obj, sum(sapply(1:3,
                          function(i) sum(apply(mix[label == i, ], 1,
                                                function(y) sqrt(sum((centroids[i,] - y)^2)))))))

}

centroids <- rbind(rep(0, 3), rep(1, 3), rep(2, 3))
label <- apply(mix, 1, classify )
label_old <- rep(1, 300)
obj = c()
while (1) {
  label <- apply(mix, 1, classify)
  if (sum(label != label_old) == 0) break
  centroids[1,] <- apply(mix[label==1,], 2, mean)
  centroids[2,] <- apply(mix[label==2,], 2, mean)
  centroids[3,] <- apply(mix[label==3,], 2, mean)
  label_old <- label
  obj = c(obj, sum(sapply(1:3,
                          function(i) sum(apply(mix[label == i, ], 1,
                                                function(y) sqrt(sum((centroids[i,] - y)^2)))))))
}
plot(obj, type = 'l')
```

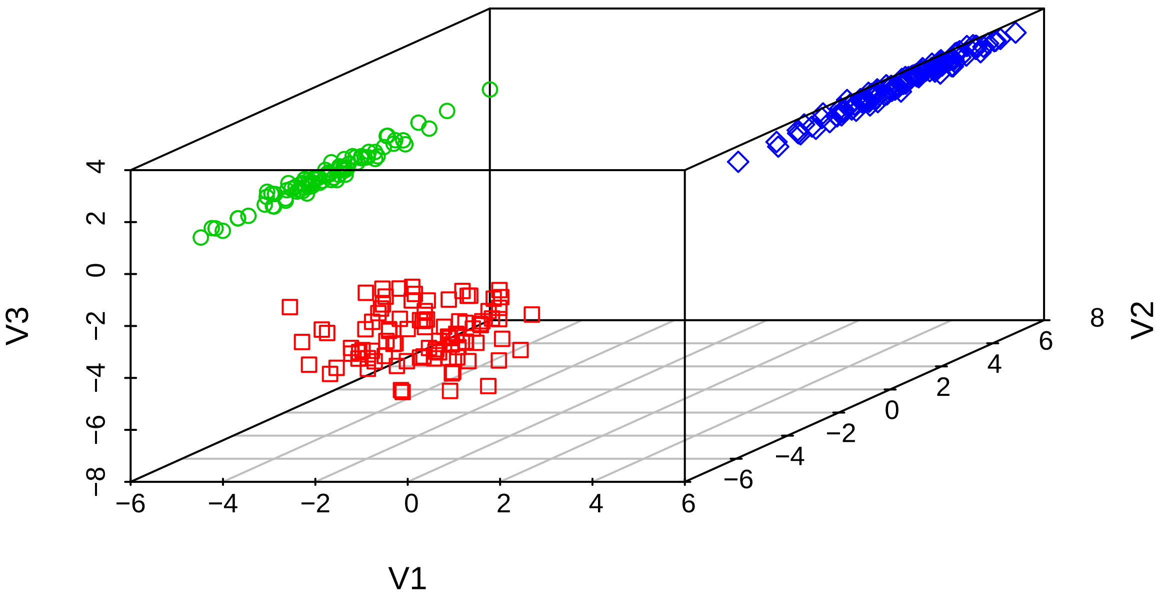


```
centroids <- rbind(rep(0,3), rep(1, 3), rep(2, 3))


pi <- rep(1/3, 3)
mu <- centroids
Sig <- array(rep(diag(3), 3), c(3, 3, 3) )

library(MASS)
library(mvtnorm)
```

```
## Warning: package 'mvtnorm' was built under R version 3.1.2
```

```
tao = matrix(0, 300, 3)
for (l in 1: 30) { # converge criterion to be added, for now just using number of iteration, works well
  for (n in 1:300) {
    fenmu <- sum(sapply(1:3, function(k) pi[k]*dmvnorm(as.numeric(mix[n,]), as.numeric(mu[k,]), Sig[,,k
    for (k in 1:3)
      tao[n,k] = pi[k]*dmvnorm(mix[n,], mu[k,], Sig[,,k])/fenmu
  }

  for (k in 1:3) {
    fenmu = sum(tao[,k])
    mu[k,] =rowSums(sapply(1:300, function(i) tao[i, k] * as.numeric(mix[i,]))) / fenmu
    Sig[,,k] = rowSums(sapply(1:300, function(i) tao[i,k] * as.numeric(mix[i,] - mu[k,])  %*% t(as.nume
    pi[k] = 1/300*fenmu
  }
}
```

```
centroids <- mu
label <- apply(mix, 1, classify )
scatterplot3d(mix[,1:3], color = label + 1, pch = ifelse(label < 3, label -1, label + 2))
```


V1
V2
V3
−6
−4
−2
0
2
4
6
8
−8
−6
−4
−2
0
2
4
−6
−4
−2
0
2
4
6