module\_11\_rmd

# Load libraries  
library(mvtnorm); library(mclust); library(e1071)  
  
# Initialize variables  
mu\_1 <- matrix(c(2,0), nrow=2)  
mu\_2 <- matrix(c(-2,0), nrow=2)  
sigma\_1 <- matrix(c(1, 0, 0, 1), nrow=2)  
P <- (1 / 2) \* matrix(c(sqrt(3), 1, -1, sqrt(3)), ncol=2)  
D <- matrix(c(9, 0, 0, 4), nrow=2)  
sigma\_2 <- P %\*% D %\*% t(P)  
  
set.seed(1) # Generate sample data  
n <- 1e3  
sample\_1 <- rmvnorm(n = n, mean = mu\_1, sigma = sigma\_1)  
sample\_2 <- rmvnorm(n = n, mean = mu\_2, sigma = sigma\_2)  
sample\_data <- rbind(sample\_1, sample\_2)  
  
# k-means  
par(mfrow = c(1,2))  
k\_mean\_clustering <- kmeans(x = sample\_data, centers = 2)  
plot(x = sample\_data[,1], y = sample\_data[,2],  
 col = c(rep('red', 1e3), rep('blue', 1e3)),  
 main = 'Sample Data Colored by Class with Centers',  
 xlab = 'x1', ylab = 'x2')  
points(k\_mean\_clustering$centers, col = 'green', pch = 19)  
  
plot(x = sample\_data[,1], y = sample\_data[,2],   
 col=k\_mean\_clustering$cluster,  
 main = 'Sample Data After k-Means Classification',  
 xlab = 'x1', ylab = 'x2')  
points(k\_mean\_clustering$centers, col = 'green', pch = 19)  
  
# Error rate 0.1975  
sum(k\_mean\_clustering$cluster != c(rep(1, 1e3), rep(2, 1e3))) / 2e3  
  
# fuzzy k-means  
fuzzy\_k\_mean\_clustering <- cmeans(x = sample\_data, centers = 2)  
plot(x = sample\_data[,1], y = sample\_data[,2],  
 col = c(rep('red', 1e3), rep('blue', 1e3)),  
 main = 'Sample Data Colored by Class with Centers',  
 xlab = 'x1', ylab = 'x2')  
points(fuzzy\_k\_mean\_clustering$centers, col = 'green', pch = 19)  
  
plot(x = sample\_data[,1], y = sample\_data[,2],  
 col=fuzzy\_k\_mean\_clustering$cluster,  
 main = 'Sample Data After Fuzzy k-Means Classification',  
 xlab = 'x1', ylab = 'x2')  
points(fuzzy\_k\_mean\_clustering$centers, col = 'green', pch = 19)  
  
# Error rate 0.193  
sum(fuzzy\_k\_mean\_clustering$cluster != c(rep(1, 1e3), rep(2, 1e3))) / 2e3  
  
# EM  
em\_clustering = Mclust(data = sample\_data, G = 2)  
plot(x = sample\_data[,1], y = sample\_data[,2],  
 col = c(rep('red', 1e3), rep('blue', 1e3)),  
 main = 'Sample Data Colored by Class with Centers',  
 xlab = 'x1', ylab = 'x2')  
points(t(em\_clustering$parameters$mean), col = 'green', pch = 19)  
plot(x = sample\_data[,1], y = sample\_data[,2],  
 col=em\_clustering$classification,  
 main = 'Sample Data After EM Classification',  
 xlab = 'x1', ylab = 'x2')  
points(t(em\_clustering$parameters$mean), col = 'green', pch = 19)  
  
# Error rate 0.1005  
sum(em\_clustering$classification != c(rep(1, 1e3), rep(2, 1e3))) / 2e3  
  
# Center comparison  
round(k\_mean\_clustering$centers, 4)  
round(fuzzy\_k\_mean\_clustering$centers, 4)  
round(t(em\_clustering$parameters$mean), 4)