Coding Assignment 3

Due Friday, March 29, 3AM

Implement the EM algorithm for a p-dimensional Gaussian mixture model with G components:

$$\sum_{k=1}^{G} p_k \cdot \mathbf{N}(x; \mu_k, \Sigma).$$

Store the parameters as a list in `R` with three components

- `prob`: a G-dimensional probability vector;

- `mean`: a p-by-G matrix with the k-th column being μ_k, the p-dimensional mean for the k-th Gaussian component;

- `Sigma`: a p-by-p covariance matrix shared by all G components.

Your code should have the following structure.

```
Estep <- function(data, G, para){
    # Your Code
    # return the n-by-G probability matrix
}

Mstep <- function(data, G, para, post.prob){
    # Your Code
    # Return the updated parameters
}

myEM <- function(data, T, G, para){
  for(t in 1:T){
      post.prob <- Estep(data, G, para)
      para <- Mstep(data, G, para, post.prob)
  }
  return(para)
}
```

You should test your code on the `faithful` data from the `R` package `mclust` with $G = 2$. The estimated parameters from your algorithm and the one from `mclust` after $T = 10$ iterations should be the same.

```r
library(mclust)
n <- nrow(faithful)
Z <- matrix(0, n, 2)
Z[sample(1:n, 120), 1] <- 1
Z[, 2] <- 1 - Z[, 1]
ini0 <- mstep(modelName="EEE", faithful, Z)$parameters

# Output from my EM alg
para0 <- list(prob = ini0$pro, mean = ini0$mean,
              Sigma = ini0$variance$Sigma)
myEM(T=10, para=para0)

# Output from mclust
Rout <- em(modelName = "EEE", data = faithful,
           control = emControl(eps=0, tol=0, itmax = 10),
           parameters = ini0)$parameters
list(Rout$pro, Rout$mean, Rout$variance$Sigma)
```

What you need to submit?

A PDF file and an R Markdown file that produces the PDF file.

- Name your files starting with

  Assignment_1_xxxx_netID

  where "**xxxx**" is the last 4-dig of your University ID.

  For example, the submission for Max Y. Chen with UID 672757127 and netID mychen12 would be named as

  Assignment_1_7127_mychen12_MaxChen.Rmd/.pdf

  You can add whatever characters after your netID.

- Your file should include the R code listed at the top of this page showing the estimated parameter from your algorithm and the one from `mclust`.

- Your file should include the derivation of the E and M steps. If you are not familiar with Latex, you do not need to include your derivation in your R Markdown, but still include your derivation in the submitted PDF file.