Coding Assignment 4

Due Monday, November 2

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

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

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

- `prob`: G-dimensional probability vector $(p_1, \ldots, p_G)$
- `mean`: p-by-G matrix with the k-th column being μ_k, the p-dimensional mean for the k-th Gaussian component;
- `Sigma`: p-by-p covariance matrix Σ 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, itmax, G, para){
  for(t in 1:itmax){
        post.prob <- Estep(data, G, para)
        para <- Mstep(data, G, para, post.prob)
      }
  return(para)
  }
```

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

```r
library(mclust)
n <- nrow(faithful)

# initialize parameters
Z <- matrix(0, n, 2)
Z[sample(1:n, 120), 1] <- 1
Z[, 2] <- 1 - Z[, 1]

ini0 <- mstep(modelName="EEE", faithful , Z)$parameters
para0 <- list(prob = ini0$pro, mean = ini0$mean,
              Sigma = ini0$variance$Sigma)

# Output from my EM alg
myEM(data=faithful, itmax=10, G=2, 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?

An R Markdown file in HTML format, which should contain all code used to produce your results.

Name your file starting with `Assignment_4_xxxx_netID` where "xxxx" is the last 4-dig of your University ID.

Your submission should include the derivation of the E- and M-steps. If you do not know how to include math formulae in R Markdown, submit the derivation as a separate PDF file. For example, write your derivation on a piece of paper, take a photo, and then save it as a PDF file; or type them in WORD and then save the file as PDF.

Name your second file, if applicable, starting with `Assignment_4_Supp_xxxx_netID` where "xxxx" is the last 4-dig of your University ID.