Computational Statistics
Chapter 3: EM algorithm

1. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ be an i.i.d. sample from a mixture of a normal distribution $\mathcal{N}(\mu, \sigma)$ and a uniform distribution $\mathcal{U}([-a, a])$, with pdf

$$f(y; \boldsymbol{\theta}) = \pi \phi(y; \mu, \sigma) + (1 - \pi)c, \tag{1}$$

where $\phi(\cdot; \mu, \sigma)$ is the normal pdf, $c = (2a)^{-1}$, π is the proportion of the normal distribution in the mixture and $\boldsymbol{\theta} = (\mu, \sigma, \pi)^T$ is the vector of parameters. Typically, the uniform distribution corresponds to outliers in the data. The proportion of outliers in the population is then $1 - \pi$. We want to estimate parameter $\boldsymbol{\theta}$ using the EM algorithm

   (a) Using the functions `sample`, `rnorm` and `runif`, generate a sample of size $n = 100$. Draw a box plot of the data.

   (b) Write an EM algorithm for this problem.

   (c) Apply the EM algorithm to the data, with different initializations. Draw the estimated probabilities $1 - z_i^{(t)}$ of being an outlier, as a function of y_i. What do you observe? Does it make sense?

2. We will now apply the same idea to linear regression. We assume that we have an independent sample $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, where the distribution of each observation Y_i a mixture of a normal distribution $\mathcal{N}(v_i^T \beta, \sigma^2)$ and a uniform distribution $\mathcal{U}([-a, a])$, v_i being a vector of covariates and β a vector of coefficients. The pdf of Y_i is

$$f(y_i; \boldsymbol{\theta}) = \pi \phi(y_i; v_i^T \beta, \sigma) + (1 - \pi)c, \tag{2}$$

with $\boldsymbol{\theta} = (\beta^T, \sigma, \pi)^T$ and $c = (2a)^{-1}$.

   (a) Using the functions `rnorm` and `runif`, generate a sample of size $n = 100$, with $v_i \sim \mathcal{U}([-6, 6])$, $\beta = (1, 2)^T$, $\sigma = 2$, $a = 20$ and $\pi = 0.5$. Draw a scatter plot of the data.

   (b) Compute the ordinary least squares estimates (OLS) of the coefficients, and draw the corresponding line.

   (c) Write an EM algorithm for this problem. (In the M-step, you will have to solve a weighted least-squares problem. You can use the `lm` function with the argument `weights`).

(d) Apply the EM algorithm to the data, taking the OLS estimates as initial values. Draw the line with coefficients equal to the MLEs. Plot the points (v_i, y_i) such that $z_i^{(t)} < 0.5$ as filled circles. Does it make sense?

3. We consider again the problem of estimating the parameters in a mixture of a normal distribution $\mathcal{N}(\mu, \sigma)$ and a uniform distribution $\mathcal{U}([-a, a])$, where a is a known constant. The observed data are an iid sample $y_1, \ldots, y_n$ from Y with pdf

$$f(y; \boldsymbol{\theta}) = \pi\phi(y; \mu, \sigma) + (1 - \pi)c, \tag{3}$$

where $\phi(\cdot; \mu, \sigma)$ is the normal pdf, $c = (2a)^{-1}$, π is the proportion of the normal distribution in the mixture and $\boldsymbol{\theta} = (\mu, \sigma, \pi)^T$ is the vector of parameters. Typically, the uniform distribution corresponds to outliers in the data. The proportion of outliers in the population is then $1 - \pi$.

We have seen how to find the MLE $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ using the EM algorithm. We now want to estimate the variance of $\widehat{\boldsymbol{\theta}}$.

(a) Set $\boldsymbol{\theta} = (0, 1, 0.9)$ and $a = 5$. Generate $N = 1000$ samples of size $n = 100$. For each sample j, compute an estimate $\widehat{\boldsymbol{\theta}}_j$ using the EM algorithm. Compute the sample variance of $\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_N$.

(b) We now consider one sample $y_1, \ldots, y_n$ and we wish to estimate $\mathrm{Var}(\widehat{\boldsymbol{\theta}})$ from that sample, without knowing the true value of $\boldsymbol{\theta}$. We will use Louis' method: compute $\widehat{i}_{\boldsymbol{x}}(\widehat{\boldsymbol{\theta}})$ and estimate $\widehat{i}_{\boldsymbol{z}|\boldsymbol{y}}(\widehat{\boldsymbol{\theta}})$ by Monte Carlo simulation; compute an estimate of $\widehat{i}_{\boldsymbol{y}}(\widehat{\boldsymbol{\theta}})$ using the missing information principle equation, and its inverse $\widehat{i}_{\boldsymbol{y}}(\widehat{\boldsymbol{\theta}})^{-1}$.