EE5111: Estimation Theory

Jan - May 2021

Mini Project 3

April 5, 2021

Study of Expectation Maximization (EM) algorithm

The objective of this exercise is to understand the Expectation Maximization (EM) algorithm. Consider the example of Gaussian mixture model as given in Example 2 of lecture 6 (page 6). Suppose $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ and W is a Bernoulli RV independent of Y_1 and Y_2 with $\pi = P(W = 0)$. An experiment is performed with n samples; for each sample a W is realized and depending on the value of W, samples are drawn from either Y_1 or Y_2. For example, when $W = 1$, the sample is drawn from Y_1 and when $W = 0$, sample is drawn from Y_2.

$$X = (1 - W)Y_1 + WY_2 \tag{1}$$

$$\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \pi) \tag{2}$$

$$f(x) = P(W = 1)f(x|W = 1) + P(W = 0)f(x|W = 0), \quad -\infty < x < \infty \tag{3}$$

$$= (1 - \pi)f_1(x) + \pi f_2(x), \quad -\infty < x < \infty \tag{4}$$

$$f_j(x) = \sigma_j^{-1}\phi\left(\frac{x - \mu_j}{\sigma_j}\right), \quad j = 1, 2 \tag{5}$$

where $\phi(z) \sim \mathcal{N}(0, 1)$.
Observed sample $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ from $f(x)$

$$l(\boldsymbol{\theta}|\boldsymbol{x}) = \sum_{i=1}^{n} \log((1 - \pi)f_1(x_i) + \pi f_2(x_i)) \tag{6}$$

In this mixture problem, unobserved data are RV which identify distribution membership. For $i = 1, 2, \cdots, n$, define

$$W_i = \begin{cases} 0, & \text{if } X_i \text{ has pdf } f_1(x) \\ 1, & \text{if } X_i \text{ has pdf } f_2(x) \end{cases} \tag{7}$$

Our aim is to obtain maximum likelihood estimate for parameter vector $\boldsymbol{\theta}$ using Expectation Maximization (EM) algorithm. Generate the observations $\boldsymbol{x}^{(i)}$ for following two cases:

(i) $\pi = 0.50$, $\mu_1 = 0.0$, $\mu_2 = 1.0$, $\sigma_1 = 0.8$ and $\sigma_2 = 0.4$

(ii) $\pi = 0.10$, $\mu_1 = 0.0$, $\mu_2 = 1.0$, $\sigma_1 = 0.8$ and $\sigma_2 = 0.4$.

Choose $n = 10, 1000, 10000$. Run the EM algorithm for following experiments.

Experiment 1: Effect of not knowing mixing fraction properly

Assume that we know π. Thus, the vector of parameters is given by $\boldsymbol{\theta} = \begin{bmatrix} \mu_1 & \mu_2 & \sigma_1 & \sigma_2 \end{bmatrix}^T$.

(a) Use observations from case (i) and assume that $\pi = 0.50$ and initial estimates are $\begin{bmatrix} 0.1 & 0.8 & 0.9 & 0.3 \end{bmatrix}^T$.

(b) Use observations from case (ii) and assume that $\pi = 0.50$ and initial estimates are $\begin{bmatrix} 0.1 & 0.8 & 0.9 & 0.3 \end{bmatrix}^T$.

(c) Use observations from case (ii) and assume that $\pi = 0.10$ and initial estimates are $\begin{bmatrix} 0.1 & 0.8 & 0.9 & 0.3 \end{bmatrix}^T$.

Experiment 2: Effect of initialization

Assume that we do not know π. Thus, the vector of parameters is given by $\boldsymbol{\theta} = \begin{bmatrix} \pi & \mu_1 & \mu_2 & \sigma_1 & \sigma_2 \end{bmatrix}^T$.

(a) Use observations from case (i) and initial estimates are $\begin{bmatrix} 0.45 & 0.1 & 0.8 & 0.9 & 0.3 \end{bmatrix}^T$.

(b) Repeat above experiment with initial estimate $\begin{bmatrix} 0.45 & 0.8 & 0.1 & 0.3 & 0.9 \end{bmatrix}^T$

(c) Use observations from case (ii) and initial estimates are $\begin{bmatrix} 0.45 & 0.1 & 0.8 & 0.9 & 0.3 \end{bmatrix}^T$.

Experiment 3: Impact of samples from distribution with closer mean

Assume that we do not know π. Thus, the vector of parameters is given by $\boldsymbol{\theta} = \begin{bmatrix} \pi & \mu_1 & \mu_2 & \sigma_1 & \sigma_2 \end{bmatrix}^T$. Now let us generate observations $\boldsymbol{x}^{(i)}$ for four different cases as follows:

(i) $\pi = 0.50$, $\mu_1 = 0.0$, $\mu_2 = 1.0$, $\sigma_1 = 0.8$ and $\sigma_2 = 0.4$

(ii) $\pi = 0.50$, $\mu_1 = 0.2$, $\mu_2 = 0.8$, $\sigma_1 = 0.8$ and $\sigma_2 = 0.4$

(iii) $\pi = 0.50$, $\mu_1 = 0.4$, $\mu_2 = 0.6$, $\sigma_1 = 0.8$ and $\sigma_2 = 0.4$

(iv) $\pi = 0.50$, $\mu_1 = 0.48$, $\mu_2 = 0.52$, $\sigma_1 = 0.8$ and $\sigma_2 = 0.4$

For all of the above cases, initial estimates are $\begin{bmatrix} 0.45 & 0.5 & 0.5 & 0.9 & 0.3 \end{bmatrix}^T$ and the number of samples in each case is $n = 5000$. What happens when two distributions are very close to each other?

Exercise

Make the following inferences from the algorithm for the aforementioned choices of n and initial estimates $\hat{\boldsymbol{\theta}}_0$:

(1) Plot the *learning curve*[1] and show the convergence of the algorithm[2] for each experiment.

(2) Observe how the final estimate of $\boldsymbol{\theta}$ from the algorithm (call it $\hat{\boldsymbol{\theta}}_{EM}$), and number of iterations needed for convergence, change when we increase n.

---

[1] Plot of the estimate at iteration $k, \hat{\theta}_k$ vs. iteration index, k.

[2] You shall consider that the algorithm has converged at iteration k when the update to any of the parameter is not more than $\epsilon = 10^{-6}$ (i.e., $\left\|\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_{k-1}\right\|_\infty = \max\left(\left|\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_{k-1}\right|\right) \leq \epsilon\right)$.

(3) Report your observation about case (b) and case (c) of experiment 1, where in former case we are assuming a wrong value of π and in later case we are assuming the true value of π.

(4) Report your observation about experiment 2(a) and 2(b) when we swap the priors μ_1 and μ_2 and priors σ_1 and σ_2.

(5) What happens when the dataset is generated with asymmetric mixing coefficients and you need to estimate that (as taken in experiment 2(c))?

(6) What happens when the means of two distributions get closer as given in experiment 3?

Submission

You are required to submit this problem no later than 19-th April 2021 (11:59 pm). Upload a compressed file containing the program/programs your four member team has written for the mini project **along with a 1-2 page report** in Moodle. The report shall include any theoretical results (if you have used any for the mini project), the final plots for each of the questions (1)-(6) and a one-two line inference on the results observed. Note that you need not include any derivations in this report. The viva for each team will be conducted jointly on a date and time convenient for all the members.