Expectation Maximisation: Exponential Mixture Model

Yukta Thapliyal, 260935739

1 Given

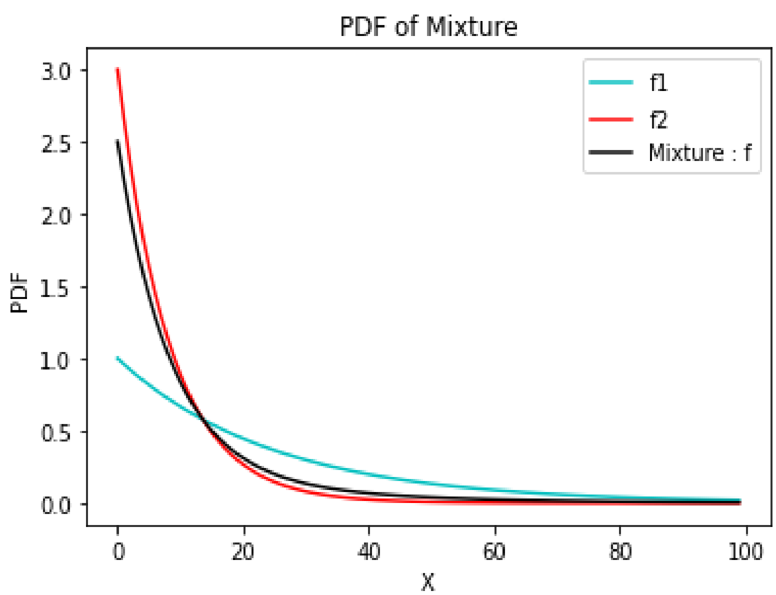
Exponential mixture models with following properties:

$$f(x;\theta) = \pi_1 f_1(x;\lambda_1) + (1-\pi_1) f_2(x;\lambda_2) \tag{1}$$

where
$\lambda_1 = 1, \lambda_2 = 3, \pi_1 = 0.25, \theta = [\lambda_1, \lambda_2, \pi_1, \pi_2]^T$

1.1 a.

Below is the pdf of exponential mixture (black) and also the pdf of original function f_1(cyan) and f_2(red).



1.2 b.

Given data, we can choose parameters to maximise the likelihood function:

$$log[f(x_1, x_2, ...x_n; \theta)] = \sum_{n=1}^{N} log f(x^n) \tag{2}$$

Let $X = (X_1, X_2, ...X_n)$ be a sample of n independent observations from mixture of two exponential distributions f_1 and f_2.

Assuming a latent variable vector Z, such that $Z = [Z_1, Z_2, ...Z_n]$. Z serves as an indicator $\mathbb{I}[Z^n = k]$, that indicates the exponential distribution k from which X_n originates i.e.

$$X_i|Z_i = 1 : f(x; \lambda_1), \quad X_i|Z_i = 0 : f(x; \lambda_2) \tag{3}$$

where

$$P(Z_i = 1) = \pi_1, \quad P(Z_i = 0) = \pi_2 = (1 - \pi_1) \tag{4}$$

The incomplete data likelihood function is:

$$L(\theta; X) = \sum_{i=n}^{n} \sum_{j=0}^{k} \pi_j \lambda_j e^{-\lambda_j X_i} \tag{5}$$

If we were to know Z_n for every X_n, i.e. the complete data likelihood function and take its logarithm, we would get:

$$\hat{\theta} = argmax_{\theta} \sum_{i=1}^{n} \sum_{j=0}^{k} I[Z_i = j][\log \pi_j + log \lambda_j - \lambda_j X_i] \tag{6}$$

We get the estimate:

$$\hat{\pi_j} = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}[Z^i = j] \tag{7}$$

$$\hat{\lambda_j} = \frac{\sum_{n=1}^{N} \mathbb{I}[Z_i = j]}{\sum_{n=1}^{N} \mathbb{I}[Z_i = j] X_i} \tag{8}$$

Where $\hat{\pi_j}$ is just our number of samples observed from j^{th} exponential distribution.

However, we do not have a complete data-set, therefore the distribution assignment $\mathbb{I}[Z_i = k]$ is unknown.

Therefore, we make use of Bayes' rule to calculate ***responsibilities*** as such:

$$P(Z = j|X) = \frac{P(X|Z = j)P(Z = j)}{P(X)} \tag{9}$$

$$\frac{\sum_{n=1}^{N} \pi_j f(X_i|\lambda_j)}{\sum_{n=1}^{N} \sum_{j=0}^{k} \pi_j f(X_i|\lambda_j)} = R_j^i \tag{10}$$

Where R^i_j tell us the probability of X_i coming from j^th exponential distribution.

Plugging in $R^i_j = P(Z_i = j|X_i)$ for $\mathbb{I}[Z_i = j]$, we have:

$$\log[L(\theta|X)] = \sum_{n=1}^{N}\sum_{j=1}^{k} R^i_j[\log[f(X_i|\lambda_j)] + \log\pi_j] \tag{11}$$

Therefore, equation 7 and 8 can now be written as:

$$\hat{\pi_j} = \frac{1}{N}\sum_{n=1}^{N} R^i_j \tag{12}$$

$$\hat{\lambda_j} = \frac{\sum_{n=1}^{N} R^i_j}{\sum_{n=1}^{N} R^i_j X_i} \tag{13}$$

EM algorithm

Since we have all pre-requisite, we can formally write EM algorithm for our given problem.

I. Expectation:

Calculate posterior probabilities from *eq* 10:

$$R^i_1 = \frac{\pi_1 f(X_i|\lambda_1)}{\pi_1 f(X_i|\lambda_1 + \pi_2 f(X_i|\lambda_2} \tag{14}$$

$$R^i_2 = \frac{\pi_2 f(X_i|\lambda_2)}{\pi_1 f(X_i|\lambda_1 + \pi_2 f(X_i|\lambda_2} \tag{15}$$

where i ranges from 1:20 in our problem

II. Maximisation:

Update *eq(s)* 13 , 14 to update θ:

$$\hat{\pi_1} = \frac{1}{N}\sum_{n=1}^{N} R^i_1 \tag{16}$$

$$\hat{\pi_2} = \frac{1}{N}\sum_{n=1}^{N} R^i_2 = 1 - \hat{\pi_1} \tag{17}$$

$$\hat{\lambda_1} = \frac{\sum_{n=1}^{N} R^i_1}{\sum_{n=1}^{N} R^i_1 X_i} \tag{18}$$

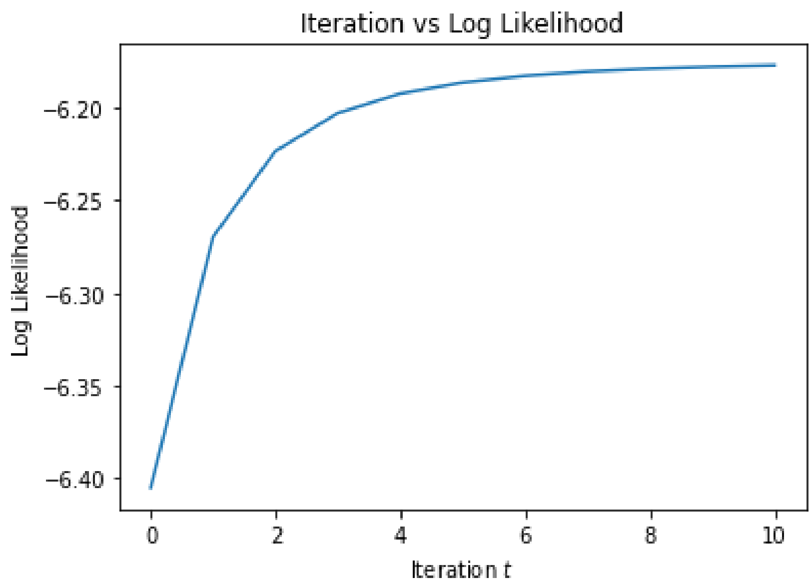
$$\hat{\lambda_2} = \frac{\sum_{n=1}^{N} R^i_2}{\sum_{n=1}^{N} R^i_2 X_i} \tag{19}$$

1.3 c.

Python code included in appendix A.

1.4 d.

Computing the log-likelihood of the incomplete set at each iteration for n = 20:
From the figure, we can clearly see that the likelihood increases monotonically



at each iteration.

For a tolerance of $1e-3$ and N = 20, we observe that it takes 11 iterations for the log-likelihood to converge with following values:

Table 1: Performance of EM with n = 20

Iteration	Log- Likelihood	Likelihood
1	-6.4053	3.9324e-7
2	-6.2697	5.3731e-7
3	-6.2234	5.9782e-7
4	-6.2029	6.2672e-7
5	-6.1923	6.4209e-7
6	-6.1864	6.5097e-7
7	-6.1828	6.5643e-7
8	-6.1804	6.5994e-7
9	-6.1789	6.6230e-7
10	-6.1778	6.6393e-7
11	-6.1771	6.6508e-7

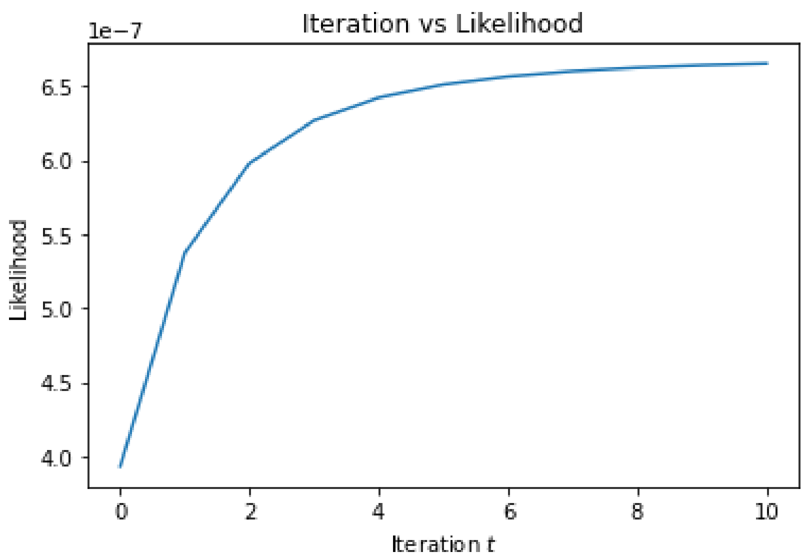
Additionally, once we vary the number of data points from n = 20 to n = 1000, we observe our θ as following:

Table 2: Performance of EM with varying n

n	π_1	π_2	λ_1	λ_2
20	0.2708	0.7291	1.4133	2.3553
200	0.2594	0.7405	1.1197	2.4861
500	0.3257	0.6742	1.1665	3.0392
1000	0.2946	0.7053	0.9693	3.5747

1.5 e.

In Table 1, we have mentioned the likelihood estimate of the exponential mixture for successive iterations. Plotting the same:



References

[1] T. Hastie et al. , The Elements of Statistical Learning, 2nd ed., Springer 2017.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin , “Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society series B, 39:1-38, 1977.