EM (Expectation-Maximization) Algorithm with An Application for Gaussian Mixture Model

Frances Lin

Fall 2020

Introduction

EM algorithm, short for “Expectation-Maximization” algorithm is an iterative method that is particular useful for finding maximum likelihood estimate (MLE) for missing data or when maximizing the likelihood function is challenging.

It is an umbrella term for a class of algorithm that iterates between expectation and maximization. Applications include EM algorithm for missing data, for censored data, for finite mixture models, etc.

Gaussian mixture model

The mixture model is made up of the incomplete (or observed) data X and the latent (or missing) variable $Z \in \{1, 2, ..., k\}$, and it is defined as

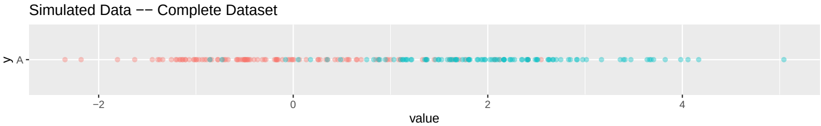
$$p(X, Z) = \sum_{k=1}^{k} p(Z)p(X|Z)$$

We cannot evaluate the (complete) likelihood because of the latent (or missing) variable Z. However, we can make use of that the posterior distribution of Z $p(Z|X)$ contains the information we need about Z.

For illustration, we apply EM algorithm to a simulated dataset that follows a two-component Gaussian mixture distribution.

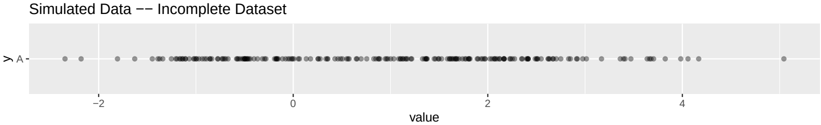
Simulated data: 2-component Gaussian mixture

We simulate 100 data points that follow $\mathcal{N}(0, 1)$ and 100 data points that follow $\mathcal{N}(2, 1)$.



Simulated data: 2-component Gaussian mixture

Now suppose that we do not know which distribution each data point is from (i.e. the latent (or missing) variable Z is involved.) This is when the EM algorithm can be of use.



EM algorithm for Gaussian mixture

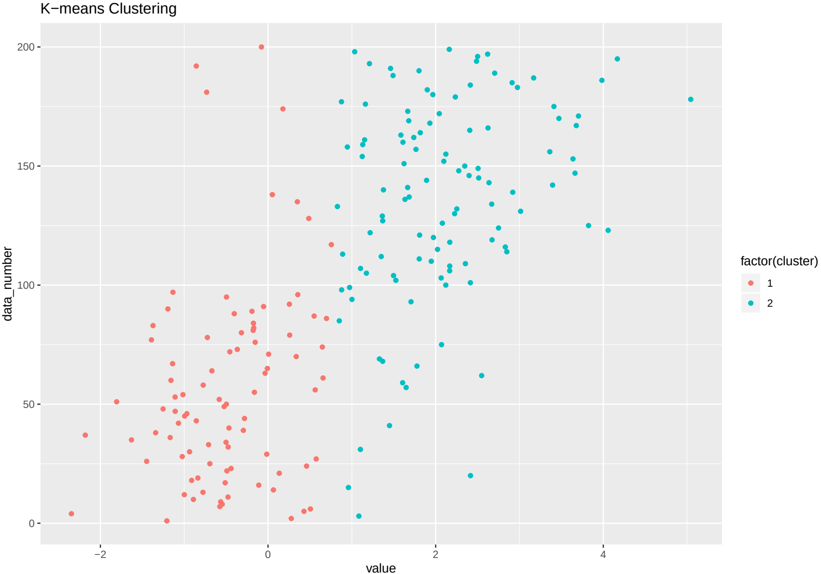
The algorithm consists of three steps:

1. Initialization
2. The E-step
3. The M-step

. . . . Iteration between E-step and M-step untill convergence

Step 1: Initialization using K-means clustering

Set initial estimates for the first EM iteration using the `kmeans` function from the `stats` package



Step 1: Initialization using K-means clustering

We obtain the estimates of parameters using MLE

$$\mu_k = \frac{\sum x_{i,k}}{N_k}, \sigma_k^2 = \frac{\sum (x_{i,k} - \mu_k)^2}{N_k}, \pi_k = \frac{N_k}{N}$$

The estimates of parameters (i.e. μ, σ, π) obtained from k-mean are:

cluster	mean	sd	pi
1	-0.4712	0.6671	0.45
2	2.099	0.8587	0.55

Step 2: The E-step

E-step in literature

$$Q(\theta, \theta^0) = E_{Z|X,\theta^0} \log(p(X, Z|\theta)) = \sum_{Z} p(Z|X, \theta^0) \log(p(X, Z|\theta))$$

Implementing

1. Calculate posterior probability (or soft labelling) $p(Z|X)$ using data X and initial estimates θ^0 and pass it to M-step
2. Compute and store the log likelihood to check for convergence

Step 2: The E-step

The output of the first 3 values of each distribution are:

value	post_1	post_2
-1.207	0.9675	0.03252
0.2774	0.4217	0.5783
1.084	0.01509	0.9849
2.415	1.54e-06	1
1.525	0.001071	0.9989
2.066	2.418e-05	1

Step 3: The M-step

M-step in literature

$$\hat{\theta} = \theta^{0+1} = argmax_{\theta} \; Q(\theta, \theta^{0})$$

Implementing

1. Replace hard labelling N_k with posterior probability (or soft labelling) $p(Z|X)$ and optimize the parameters using MLE
2. Return the final estimates that maximize the likelihood, if convergence happens

Step 3: The M-step

After replacing, we have that

$$\mu_k = \frac{\sum_{i=1}^{N} p(Z_i|X_i)x_i}{\sum_{i=1}^{N} p(Z_i|X_i)}, \sigma_k^2 = \frac{\sum_{i=1}^{N} p(Z_i|X_i)(x_i - \mu_k)^2}{\sum_{i=1}^{N} p(Z_i|X_i)}$$

$$\pi_k = \frac{\sum_{i=1}^{N} p(Z_i|X_i)}{N}$$

The final estimates of parameters (i.e. μ, σ, π) are:

cluster	mean	sd	pi
1	-0.6611	0.591	0.3447
2	1.786	1.078	0.6553

Convergence: Iterate between E-step and M-step untill convergence

Compare log-likelihood of current iteration to log-likelihood of previous iteraction to see if the change is minimal

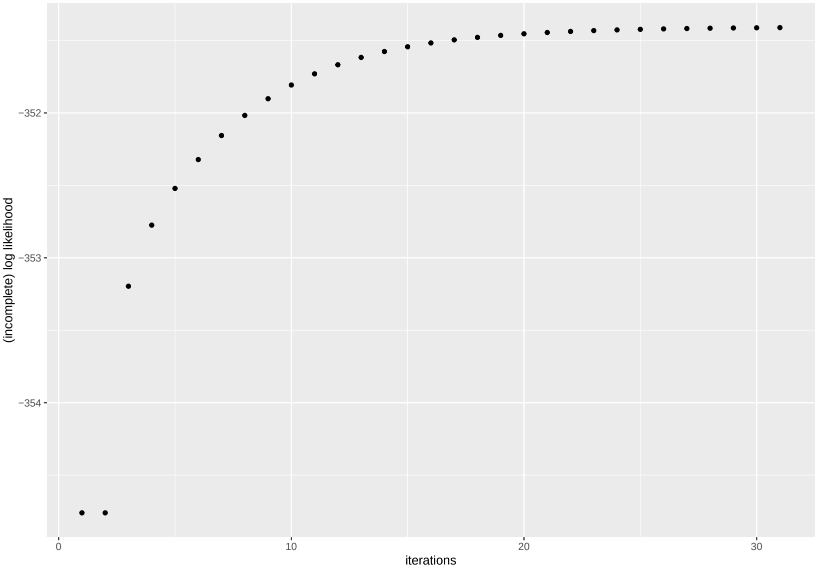
If the change is minimal (i.e. convergence), we stop and return the final estimates. If it isn't, then we repeat another EM step.

```
...

...

if (convergence){

 break

} else {

 1. set log-likelihood$^{0+1}$ to log-likelihood$^{0}$

 2. repeat E-step and M-step

}
```

Convergence: Iterate between E-step and M-step untill convergence

max_log_likelihood	#s of iterations
-351.4	31

Convergence: Iterate between E-step and M-step untill convergence



Thank you!

EM (Expectation-Maximization) Algorithm with An Application for Gaussian Mixture Model

Frances Lin

MS student, Dept. of Statistics, Oregon State University

GitHub project repository