Computational Statistics Midterm: EM for mixture distribution

Pei-Shan Yen, Jieqi Tu, Jun Lu, Hajwa Kim, Yanli Gao

Nov 25 2020

1 Introduction

The random variable Y, finite mixture distribution, defined as a random drawn from a collection of underlying individual random variables $X_j, j = 1, 2, ..., k$ with the corresponding probability density function f_j. The individual random variables that are combined to form the mixture distribution are called the mixture components, and the probabilities (or weights) associated with each component are called the mixture weights. Assume the chance that the mixture Y following X_j is the mixture weight p_j. The sum of the mixture weights should be 1, i.e., $\sum_{j=1}^{k} p_j = 1$.

The probability density function of the mixture Y is

$$f(y) = \sum_{j=1}^{k} p_j f_j \tag{1}$$

If a mixture dataset with sample size n is observed, and the ith observation is from the jth component, the indicator variable δ can be defined as

$$\delta_{ij} = 1 \ when \ y_i \sim f_j \tag{2}$$

Obviously, $\underset{\sim}{\delta_i} = (\delta_{i1}, \delta_{i2}, ..., \delta_{ik})$ follows a multinomial distribution

$$\underset{\sim}{\delta_i} \sim MN(size = 1; p_1, p_2, .., p_k) \tag{3}$$

2 EM algorithm

2.1 the mixture of Poisson distribution

Assume the underlying individual distribution is Poisson distribution with rate parameter λ_j. We have

$$Y_i | \underset{\sim}{\delta_i} \sim Poi(\lambda_j) \tag{4}$$

The joint distribution of (Y, δ) is

$$f(\underset{\sim}{Y}, \underset{\sim}{\delta}) = \prod_{i=1}^{n} f(\underset{\sim}{Y_i} | \underset{\sim}{\delta_i}) f(\underset{\sim}{\delta_i}) = \prod_{i=1}^{n} \prod_{j=1}^{k} \Big[(e^{-\lambda_j} \frac{\lambda_j^{y_i}}{y_i!}) p_j \Big]^{\delta_{ij}} \tag{5}$$

The log-likelihood of (Y, δ) is

$$log f(\underset{\sim}{Y}, \underset{\sim}{\delta}) = \sum_{i=1}^{n} \sum_{j=1}^{k} \delta_{ij} \Big[- \lambda_j + y_i log \lambda_j - y_i! + log p_j \Big] \tag{6}$$

The Expectation-Maximization algorithm (EM) can be used to estimate the unknown parameter $\theta = (p_1, p_2, ..., p_{k-1}, \lambda_1, \lambda_2, ..., \lambda_k)$ for the mixture distribution. Given the initial value $\theta^{(0)}$, the objective function Q is defined as

$$Q(\theta | \theta^{(0)}, \underset{\sim}{Y}) = E_{\delta_{ij}} \Big[log f(\underset{\sim}{Y}, \underset{\sim}{\delta}) \Big] = \sum_{i=1}^{n} \sum_{j=1}^{k} E \Big[\delta_{ij} \Big] \Big[- \lambda_j + y_i log \lambda_j - y_i! + log p_j \Big] \tag{7}$$

From equation (7), the conditional expectation δ_{ij} given the observation Y_i at iteration t of the algorithm is

$$E[\delta_{ij}|Y_i] = P(\delta_{ij}=1|Y_i) = \frac{P(Y_i|\delta_{ij}=1)P(\delta_{ij}=1)}{\sum\limits_{j=1}^{k} P(Y_i|\delta_{ij}=1)P(\delta_{ij}=1)} = \frac{f_j^{(t)}p_j^{(t)}}{\sum\limits_{m=1}^{k} f_m^{(t)}p_m^{(t)}} = \frac{\left[e^{-\lambda_j^{(t)}}\frac{\lambda_j^{(t)y_i}}{y_i!}\right]p_j^{(t)}}{\sum\limits_{m=1}^{k}\left[e^{-\lambda_m^{(t)}}\frac{\lambda_m^{(t)y_i}}{y_i!}\right]p_m^{(t)}} = w_{ij}^{(t)} \tag{8}$$

Hence, in the E-step, the objective function Q can be simplified to

$$Q(\theta|\theta^{(0)}, \underset{\sim}{Y}) = Q(p_j, \lambda_j|p_j^{(t)}, \lambda_j^{(t)}, \underset{\sim}{Y}) = \sum_{i=1}^{n}\sum_{j=1}^{k} w_{ij}^{(t)}(-\lambda_j + y_i ln\lambda_j - y_i! + logp_j) \tag{9}$$

In the M-step, we maximize the objective function Q with respect to θ.

For the Poisson rate parameter λ_j, we have

$$\frac{dQ}{d\lambda_j} = \frac{dE_{\delta_{ij}}\left[logf(\underset{\sim}{Y}, \underset{\sim}{\delta})\right]}{d\lambda_j} = w_{ij}^{(t)}\left(\frac{y_i}{\lambda_j} - 1\right) \tag{10}$$

The maximum likelihood estimate of $\lambda_j^{(t+1)}$ is

$$\lambda_j^{(t+1)} = \frac{\sum\limits_{i=1}^{n} w_{ij}^{(t)} y_i}{\sum\limits_{i=1}^{n} w_{ij}^{(t)}} \tag{11}$$

For the mixture weight parameter p_j, we have

$$\frac{dQ}{dp_j} = \frac{dE_{\delta_{ij}}\left[logf(\underset{\sim}{Y}, \underset{\sim}{\delta})\right]}{dp_j} = \sum_{i=1}^{n}\left[\frac{w_{ij}^{(t)}}{p_j} - \frac{w_{ik}^{(t)}}{p_k}\right] \tag{12}$$

The maximum likelihood estimate of $p_j^{(t+1)}$ is

$$p_j^{(t+1)} = \frac{w_{ij}^{(t)}}{n} \tag{13}$$

2.2 the mixture of Exponential distribution

Assume the underlying individual distribution is Exponential distribution with rate parameter λ_j. We have

$$Y_i|\underset{\sim}{\delta_i} \sim Exp(\lambda_j) \tag{14}$$

The joint distribution of (Y, δ) is

$$f(\underset{\sim}{Y}, \underset{\sim}{\delta}) = \prod_{i=1}^{n} f(\underset{\sim}{Y_i}|\underset{\sim}{\delta_i})f(\underset{\sim}{\delta_i}) = \prod_{i=1}^{n}\prod_{j=1}^{k}\left[\left(\lambda_j e^{-\lambda_j y_i}\right)p_j\right]^{\delta_{ij}} \tag{15}$$

The log-likelihood of (Y, δ) is

$$logf(\underset{\sim}{Y}, \underset{\sim}{\delta}) = \sum_{i=1}^{n}\sum_{j=1}^{k}\delta_{ij}\left[log\lambda_j - y_i\lambda_j + logp_j\right] \tag{16}$$

In E-step, the objective function Q is

$$Q(p_j, \lambda_j|p_j^{(t)}, \lambda_j^{(t)}, \underset{\sim}{Y}) = \sum_{i=1}^{n}\sum_{j=1}^{k} w_{ij}^{(t)}(log\lambda_j - y_i\lambda_j + logp_j) \tag{17}$$

where

$$w_{ij}^{(t)} = \frac{f_j^{(t)}p_j^{(t)}}{\sum\limits_{m=1}^{k} f_m^{(t)}p_m^{(t)}} = \frac{\left[\lambda_j^{(t)}e^{-\lambda_j^{(t)}y_i}\right]p_j^{(t)}}{\sum\limits_{m=1}^{k}\left[\lambda_m e^{-\lambda_m^{(t)}y_i}\right]p_m^{(t)}} \tag{18}$$

In M-step, the maximum likelihood estimate of $p_j^{(t+1)}$ is identical to equations (13). The maximum likelihood estimate of $\lambda_j^{(t+1)}$ is $\frac{\sum_{i=1}^{n} w_{ij}^{(t)}}{\sum_{i=1}^{n} w_{ij}^{(t)}y_i}$.

2.3 the mixture of Rayleigh distribution

Assume the underlying individual distribution is Rayleigh distribution with parameter σ_j. We have

$$Y_i|\underset{\sim}{\delta}_i \sim Rayleigh(\sigma_j) \tag{19}$$

The joint distribution of (Y, δ) is

$$f(\underset{\sim}{Y}, \underset{\sim}{\delta}) = \prod_{i=1}^{n} f(\underset{\sim}{Y}_i|\underset{\sim}{\delta}_i) f(\underset{\sim}{\delta}_i) = \prod_{i=1}^{n}\prod_{j=1}^{k}\left[(\frac{y_i}{\sigma_j^2} e^{\frac{-y_i^2}{2\sigma_j^2}})p_j\right]^{\delta_{ij}} \tag{20}$$

The log-likelihood of (Y, δ) is

$$logf(\underset{\sim}{Y}, \underset{\sim}{\delta}) = \sum_{i=1}^{n}\sum_{j=1}^{k}\delta_{ij}\left[-\frac{y_i^2}{2\sigma_j^2} + logy_i - 2log\sigma_j + logp_j\right] \tag{21}$$

In E-step, the objective function Q is

$$Q(p_j, \lambda_j|p_j^{(t)}, \lambda_j^{(t)}, \underset{\sim}{Y}) = \sum_{i=1}^{n}\sum_{j=1}^{k} w_{ij}^{(t)}(-\frac{y_i^2}{2\sigma_j^2} + logy_i - 2log\sigma_j + logp_j) \tag{22}$$

where

$$w_{ij}^{(t)} = \frac{f_j^{(t)}p_j^{(t)}}{\sum\limits_{m=1}^{k} f_m^{(t)}p_m^{(t)}} = \frac{\left[\frac{y_i}{\sigma_j^2(t)} e^{\frac{-y_i^2}{2\sigma_j^2(t)}}\right]p_j^{(t)}}{\left[\sum\limits_{m=1}^{k}\frac{y_i}{\sigma_m^2(t)} e^{\frac{-y_i^2}{2\sigma_m^2(t)}}\right]p_m^{(t)}} \tag{23}$$

In M-step, the maximum likelihood estimate of $p_j^{(t+1)}$ is identical to equations (13). The maximum likelihood estimate of $\sigma_j^{(t+1)}$ is $\sqrt{\frac{\sum_{i=1}^{n} w_{ij}^{(t)} y_i^2}{2\sum_{i=1}^{n} w_{ij}^{(t)}}}$.

The detailed EM algorithm for each underlying distribution are included in Appendix.