Expectation Maximization – An Example

Amro Al Baali

November 29, 2020

1 Notations

Let the observations be $X = \{x_1, \ldots, x_n\}$ be the set of realizations of the random variable $\underline{x} \in \mathbb{R}$. Additionally, let $Z = \{z_1, \ldots, z_n\}$ be the set of realizations of $\underline{z} \in \{0, 1\}$ which is the "missing" data of the problem.[1] The set of complete data is denoted by $Y = X, Z$. Furthermore, let the two PDF that $\underline{x}$ is distributed from be denoted by $f_1(x; \boldsymbol{\lambda}_1)$ and $f_2(x; \boldsymbol{\lambda}_2)$, where $\boldsymbol{\lambda}_i$ is the set of parameters for each of the two PDFs.

[1] The data is "missing" in the sense that, if this data was available, then the problem would be significantly simplified.

The set of unknown parameters are $\boldsymbol{\theta} = \{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \pi_1\}$.

2 Problem statement

Let $\underline{x} \in \mathbb{R}$ be a random variable[2] that can be sampled from one of two distributions $f_1(x; \boldsymbol{\lambda}_1)$ and $f_2(x; \boldsymbol{\lambda}_2)$.[3] Whether $\underline{x}$ will be sampled from $f_1(x; \boldsymbol{\lambda}_1)$ or $f_2(x; \boldsymbol{\lambda}_2)$ will depend on another random variable $\underline{z} \in \{0, 1\}$. Specifically, the conditional PDF of $\underline{x}$ given $\underline{z}$ is

[2] In this document a single random variable will be used. It is possible to generalization to multivariate random variables.

[3] In this document, we'll assume that there are only two possible distributions but the idea generalizes to multiple distributions.

$$f(x \mid z; \boldsymbol{\lambda}) = \begin{cases} f_1(x; \boldsymbol{\lambda}_1), & z = 0, \\ f_2(x; \boldsymbol{\lambda}_2), & z = 1, \end{cases} \tag{1}$$

where $\boldsymbol{\lambda} = \{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2\}$. Let the PMF of $\underline{z}$ be given by

$$p(z) = \begin{cases} \pi_1, & z = 0, \\ 1 - \pi_1, & z = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The problem is to estimate the parameters $\boldsymbol{\theta}$ without having known π_1. The set of unknown parameters will be denoted by $\boldsymbol{\theta} = \{\boldsymbol{\lambda}, \pi_1\}$.

3 Doing "the math"

The PDF $f(y; \boldsymbol{\theta})$ is needed in the expectation step. The PDF is given by

$$\begin{aligned} f(y; \boldsymbol{\theta}) &= f(x, z; \boldsymbol{\theta}) & (3) \\ &= f(x \mid z; \boldsymbol{\theta}) f(z; \boldsymbol{\theta}). & (4) \end{aligned}$$

Plugging (??) and (??) into (??) gives

$$f(x, z; \boldsymbol{\theta}) = f(x \mid z; \boldsymbol{\theta}) f(z; \boldsymbol{\theta}) \tag{5}$$

$$= \begin{cases} f_1(x; \boldsymbol{\lambda}_1) \pi_1, & z = 0, \\ f_2(x; \boldsymbol{\lambda}_2)(1 - \pi_1), & z = 1, \end{cases} \tag{6}$$

which can be rewritten[4] as

$$f(x, z; \boldsymbol{\theta}) = (\pi_1 f_1(x; \boldsymbol{\lambda}_1))^{1-z} ((1 - \pi_1) f_2(x; \boldsymbol{\lambda}_2))^z . \tag{7}$$

[4] This may be confusing at first, by simply replace z with 0 or 1 and the expression (??) will be exactly recovered.

The marginal PDF on $\underline{x}$ is obtained by marginalizing out $\underline{z}$ from $f(x, z; \boldsymbol{\theta})$ to give

$$f(x; \boldsymbol{\theta}) = \sum_{i=0}^{1} f(x, z = i; \boldsymbol{\theta}) \tag{8}$$

$$= \pi_1 f_1(x; \boldsymbol{\theta}) + (1 - \pi_1) f_2(x; \boldsymbol{\theta}) \tag{9}$$

$$= \pi_1 f_1(x; \boldsymbol{\lambda}_1) + (1 - \pi_1) f_2(x; \boldsymbol{\lambda}_2) . \tag{10}$$

4 *The expectation step*

From [5] and [6], the function to be maximized is $Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right)$,[7] which is the expectation of the log-likelihood of the complete data. That is,

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}) = \mathbb{E}_{\underline{Y}} \left[\log f(\underline{Y}; \boldsymbol{\theta}) \mid X, \boldsymbol{\theta}^{(j)} \right] . \tag{11}$$

[5]

[6]

[7] $\boldsymbol{\theta}^{(j)}$ is the jth estimate of $\boldsymbol{\theta}$.

The log-likelihood function of the complete data is given by

$$\log f(\underline{Y}; \boldsymbol{\theta}) = \sum_{i=1}^{n} \log \Big((\pi_1 f_1(\underline{x}_i; \boldsymbol{\lambda}_1))^{1-\underline{z}_i} \cdot ((1 - \pi_1) f_2(\underline{x}_i; \boldsymbol{\lambda}_2))^{\underline{z}_i} \Big) \tag{12}$$

$$= \sum_{i=1}^{n} (1 - \underline{z}_i)(\log \pi_1 + \log f_1(\underline{x}_i; \boldsymbol{\lambda}_1)) + \sum_{i=1}^{n} \underline{z}_i (\log(1 - \pi_1) + \log f_2(\underline{x}_i; \boldsymbol{\lambda}_2)) \tag{13}$$

$$= \sum_{i=1}^{n} (1 - \underline{z}_i) \log f_1(\underline{x}_i; \boldsymbol{\lambda}_1) + \underline{z}_i \log f_2(\underline{x}_i; \boldsymbol{\lambda}_2) + \sum_{i=1}^{n} \underline{z}_i \log(1 - \pi_1) + (1 - \underline{z}_i) \log \pi_1 . \tag{14}$$

The conditional expectation (**??**) can then be expanded to give

$$
\begin{aligned}
Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right) &= \mathbb{E}_{\underline{Y}}\left[\log f\left(\underline{Y};\boldsymbol{\theta}\right) \mid X, \boldsymbol{\theta}^{(j)}\right] && (15)\\
&= \mathbb{E}_{\underline{Z}}\left[\log f\left(\underline{X},\underline{Z};\boldsymbol{\theta}\right) \mid \underline{X} = X, \boldsymbol{\theta}^{(j)}\right] && (16)\\
&= \sum_{i=1}^{n}\left(1-\mathbb{E}\left[\underline{z}_i \mid X,\boldsymbol{\theta}^{(j)}\right]\right)\log f_1\left(x_i;\boldsymbol{\lambda}_1^{(j)}\right) + \mathbb{E}\left[\underline{z}_i \mid X,\boldsymbol{\theta}^{(j)}\right]\log f_2\left(x_i;\boldsymbol{\lambda}_2^{(j)}\right)\\
&\quad + \sum_{i=1}^{n}\mathbb{E}\left[\underline{z}_i \mid X,\boldsymbol{\theta}^{(j)}\right]\log \pi_1^{(j)} + (1-\mathbb{E}\left[\underline{z}_i \mid X,\boldsymbol{\theta}^{(j)}\right])\log\left(1-\pi_1^{(j)}\right) && (17)\\
&= \sum_{i=1}^{n}\mathbb{E}\left[\underline{z}_i \mid X,\boldsymbol{\theta}^{(j)}\right]\left(\log f_2\left(x_i;\boldsymbol{\lambda}_2^{(j)}\right) - \log f_1\left(x_i;\boldsymbol{\lambda}_1^{(j)}\right)\right)\\
&\quad + \sum_{i=1}^{n}\left(\log \pi_1^{(j)} - \log\left(1-\pi_1^{(j)}\right)\right)\\
&\quad + \sum_{i=1}^{n}\log f_1\left(x_i;\boldsymbol{\lambda}_1^{(j)}\right) + \log\left(1-\pi_1^{(j)}\right). && (18)
\end{aligned}
$$

The expectation $\mathbb{E}\left[\underline{z}_i \mid X,\boldsymbol{\theta}^{(j)}\right]$ is simplified to $\mathbb{E}_{z_i}\left[\underline{z}_i \mid x_i,\boldsymbol{\theta}^{(j)}\right]$ since

$$
\begin{aligned}
f\left(z_i \mid X\right) &= \frac{f\left(X \mid z_i\right) f\left(z_i\right)}{f\left(X\right)} && (19)\\
&= \frac{f\left(x_1\right)\cdots f\left(x_i \mid z_i\right)\cdots f\left(x_n\right) f\left(z_i\right)}{f\left(x_1\right)\cdots f\left(x_i\right)\cdots f\left(x_n\right)} && (20)\\
&= \frac{f\left(x_i \mid z_i\right) f\left(z_i\right)}{f\left(x_i\right)} && (21)\\
&= f\left(z_i \mid x_i\right), && (22)
\end{aligned}
$$

where the independence of $\underline{x}_j$ from $\underline{z}_i$ for $i \neq j$ was used. The expectation is therefore

$$
\begin{aligned}
\mathbb{E}\left[\underline{z}_i \mid x_i,\boldsymbol{\theta}^{(j)}\right] &= \sum_{i=0}^{1}\frac{f\left(x,z;\boldsymbol{\theta}\right)}{f\left(x;\boldsymbol{\theta}\right)} z_i && (23)\\
&= \frac{1}{f\left(x;\boldsymbol{\theta}\right)}\sum_{i=0}^{1} f\left(x,z;\boldsymbol{\theta}\right) z_i && (24)\\
&= \frac{1}{f\left(x;\boldsymbol{\theta}^{(j)}\right)}\left(f_1\left(x;\boldsymbol{\lambda}_1^{(j)}\right)\pi_1^{(j)}(0)\right.\\
&\qquad \left. + f_2\left(x;\boldsymbol{\lambda}_2^{(j)}\right)(1-\pi_1^{(j)})(1)\right) && (25)\\
&= \frac{f_2\left(x;\boldsymbol{\lambda}_2^{(j)}\right)\left(1-\pi_1^{(j)}\right)}{\pi_1^{(j)} f_1\left(x;\boldsymbol{\lambda}^{(j)}\right) + \left(1-\pi_1^{(j)}\right) f_2\left(x;\boldsymbol{\lambda}_2^{(j)}\right)}. && (26)
\end{aligned}
$$