EM-example : Discrete Mixture model.

Suppose we have a Mixture model with pmf $p(x_i) = \gamma p_1(x_i) + (1-\gamma)p_2(x_i)$

Where x_i can take in the values $\{1,2,3\}$. We have the following values:

distribution	$p(x_i = 1 \vert t_i = c)$	$p(x_i = 2 \vert t_i = c)$	$p(x_i = 3 \vert t_i = c)$
$c=1$	α	$1-\alpha$	0
$c=2$	0	$1-\beta$	β

The goal is to find out the parameters α, β, γ by using the EM algorithm. Suppose we set

$\alpha_0 = \beta_0 = \gamma_0 = 0.5$ for an initialization.

E-step

First define the latent variable t_i that influences x_i. Now, think of t_i as the 'group label'. A variable t_i means that it is from distribution p_i. From above, t_i can only take in values $\{1,2\}$

Hence, $P(t_i = 1) = \gamma$, and $p(x_i|t_i = j) = p_j(x_i)$.

Finally, Let's start with the E-step.

$q(t_i = c) = p(t_i = c|x_i)$ (x_i's latent variable distribution is based on the RHS conditional)

So let's start with $p(t_i = 1|x_i = 1)$. By bayes rule, we have:

$$
\begin{aligned}
p(t_i = 1|x_i = 1) &= \frac{p(t_i = 1, x_i = 1)}{p(x_i = 1)} && (1) \\
&= \frac{p(x_i = 1|t_i = 1)p(t_i = 1)}{p(x_i = 1|t_i = 1)p(t_i = 1) + p(x_i = 1|t_i = 2)p(t_i = 2)} && (2) \\
&= \frac{\alpha \cdot \gamma}{\alpha \cdot \gamma + o \cdot (1-\gamma)} = 1 && (3)
\end{aligned}
$$

This is certain. If we get a $x_i = 1$, we can be certain that it was generated by latent variable $t_i = 1$.

We can compute other things as well:

$$
\begin{aligned}
p(t_i = 1|x_i = 2) &= \frac{p(t_i = 1, x_i = 2)}{p(x_i = 2)} && (4) \\
&= \frac{p(x_i = 2|t_i = 1)p(t_i = 1)}{p(x_i = 2|t_i = 1)p(t_i = 1) + p(x_i = 2|t_i = 2)p(t_i = 2)} && (5) \\
&= \frac{(1-\alpha)\gamma}{(1-\alpha)\gamma + (1-\beta)(1-\gamma)} = 0.5 \text{ (by using initial values)} && (6)
\end{aligned}
$$

$p(t_i = 1|x_i = 3) = 0$ because p_1 has probability 0 for $p_1(x_i = 3)$.

Summarizing, we obtain that:

$$q(t_i = 1) = p(t_i = 1|x_i) = \begin{cases} 1, \ x_i = 1 \\ 0.5, \ x_i = 2 \\ 0, \ x_i = 3 \end{cases}$$

And

$q(t_i = 2) = p(t_i = 2|x_i) = 1 - q(t_i = 1)$.

We found the q distributions now. It is time to do the M-step.

M-step

Suppose we observed that $N_1 = 30$, $N_2 = 20$, $N_3 = 60$ in a dataset of 110, where N_i is the count for variables that outputted $i \in \{1, 2, 3\}$.

The objective is to obtain

$$\underset{\alpha,\beta,\gamma}{\operatorname{argmax}} \sum_{i=1}^{N} \mathbb{E}_{q(t_i)} \left[\log p(x_i|t_i) p(t_i) \right] \tag{7}$$

(NOTE: we do the product $p(x_i|t_i)p(t_i)$ because we have to account for the variables $\theta = \{\alpha, \beta, \gamma\}$. So it actually is : $p(x_i|t_i, \theta)p(t_i|\theta) = p(x_i, t_i|\theta)$, which is used for the M step of the EM algorithm).

Now,

$$\begin{align} \sum_{i=1}^{N} \mathbb{E}_{q(t_i)} \left[\log p(x_i|t_i) p(t_i) \right] &= \sum_{i=1}^{N} q(t_i = 1) \log p(x_i|t_i = 1)\gamma + \sum_{i=1}^{N} q(t_i = 2) \log p(x_i|t_i = 2)(1-\gamma) \tag{8} \\ &= N_1 p(t_i = 1|x_i = 1) \log(\alpha)\gamma + N_2 p(t_i = 1|x_i = 2) log(1-\alpha)\gamma \tag{9} \\ &+ N_3 \underbrace{p(t_i = 1|x_i = 3)}_{\text{(this is 0)}} log(0)\gamma \tag{10} \\ &+ N_1 \underbrace{p(t_i = 2|x_i = 1)}_{\text{(this is 0)}} \log(0)(1-\gamma) + N_2 p(t_i = 2|x_i = 2) log(1-\beta)(1-\gamma) \tag{11} \\ &+ N_3 p(t_i = 2|x_i = 3) log(\beta)(1-\gamma) \tag{12} \\ &= N_1 \alpha\gamma \log\alpha + 0.5 N_2 \gamma \log(1-\alpha) + N_2 0.5(1-\gamma)\log(1-\beta) + N_3 log(\beta)(1-\gamma) \tag{13} \end{align}$$

By taking partial derivatives $\frac{\partial}{\partial\alpha}, \frac{\partial}{\partial\beta}, \frac{\partial}{\partial\gamma}$, of the above, and setting to 0 for each, we obtain $\alpha = 0.75, \beta = \frac{6}{7}, \gamma = \frac{4}{11}$.