Geometric EM

You are given a mixture model with mixture components c_1, c_2 which are linked to geometric distributions with parameters $\theta_{c_1}^{(0)} = 0.2, \theta_{c_2}^{(0)} = 0.6$. You observe the data set

$$\{0, 2, 2, 3\} \ .$$

Assume that the latent variables are i.i.d. and that $P(Y = c_1 | \Theta = \theta^{(0)}) = 0.2$.

a) What is the (marginal) log-likelihood of this data set under the model? Feel free to use calculators.

b) Find the most likely mixture component for each data point.

c) Perform one EM iteration.

d) Compute the marginal log-likelihood of the data with the updated parameters. The new value should be higher than the one computed in the beginning.

Geometric EM

a) We observed data $x_1, x_2, x_3, x_4 = 0, 2, 2, 3$; and have initial parameters $\theta^{(0)} = (\theta_{c_1}^{(0)}, \theta_{c_2}^{(0)}, w_1^{(0)}, w_2^{(0)}) = (0.2, 0.6, 0.2, 0.8)$. Therefore, the log-likelihood is:

$$
\begin{aligned}
& \log P(X = x|\Theta = \theta^{(0)}) \\
= & \log P(X_1 = x_1|\Theta = \theta^{(0)}) + \log P(X_2 = x_2|\Theta = \theta^{(0)}) \\
& \quad + \log P(X_3 = x_3|\Theta = \theta^{(0)}) + \log P(X_4 = x_4|\Theta = \theta^{(0)}) \\
= & \log\left[P(X_1 = 0, Y_1 = c_1|\Theta = \theta^{(0)}) + P(X_1 = 0, Y_1 = c_2|\Theta = \theta^{(0)})\right] \\
& \quad + \log P(X_2 = 2|\Theta = \theta^{(0)}) + \log P(X_3 = 2|\Theta = \theta^{(0)}) \\
& \quad + \log P(X_4 = 3|\Theta = \theta^{(0)}) \\
= & \log(0.2 \cdot 0.2 \cdot 0.8^0 + 0.8 \cdot 0.6 \cdot 0.4^0) \\
& \quad + 2 \cdot \log(0.2 \cdot 0.2 \cdot 0.8^2 + 0.8 \cdot 0.6 \cdot 0.4^2) \\
& \quad + \log(0.2 \cdot 0.2 \cdot 0.8^3 + 0.8 \cdot 0.6 \cdot 0.4^3) = -8.1836793
\end{aligned}
$$

Geometric EM

b) $P(Y = c_1|X = 0, \Theta = \theta^{(0)}) = \dfrac{0.2 \cdot 0.2 \cdot 0.8^0}{0.2 \cdot 0.2 \cdot 0.8^0 + 0.8 \cdot 0.6 \cdot 0.4^0} = 0.0769231.$
Also $P(Y = c_2|X = 0, \Theta = \theta^{(0)}) = 1 - 0.0769231 = 0.9230769.$
Therefore, c_2 is the more likely mixture component when observing $X = 0$.

$$P(Y = c_1|X = 2, \Theta = \theta^{(0)}) = \frac{0.2 \cdot 0.2 \cdot 0.8^2}{0.2 \cdot 0.2 \cdot 0.8^2 + 0.8 \cdot 0.6 \cdot 0.4^2} = 0.25$$

Also $P(Y = c_2|X = 2, \Theta = \theta^{(0)}) = 1 - 0.25 = 0.75$. Therefore, c_2 is the more likely mixture component when observing $X = 2$.

$$P(Y = c_1|X = 3, \Theta = \theta^{(0)}) = \frac{0.2 \cdot 0.2 \cdot 0.8^3}{0.2 \cdot 0.2 \cdot 0.8^3 + 0.8 \cdot 0.6 \cdot 0.4^3} = 0.4$$

Also $P(Y = c_2|X = 3, \Theta = \theta^{(0)}) = 1 - 0.4 = 0.6$. Therefore, c_2 is the more likely mixture component when observing $X = 3$.

Go play with the applet in order to check that this makes sense!

Geometric EM

c) **E-step**

expected fractional count of c_1:

$$\mathbb{E}[\sum_i \mathbb{1}\left(Y_i = c_1\right) \mid X = x, \Theta = \theta^{(0)}] = 0.0769231 + 2 \cdot 0.25 + 0.4$$

$$= 0.9769231$$

expected fractional count of c_2:

$$\mathbb{E}[\sum_i \mathbb{1}\left(Y_i = c_2\right) \mid X = x, \Theta = \theta^{(0)}] = 4 - 0.9769231 = 3.0230769$$

expected sufficient statistic $\sum_i x_i$:

$$\mathbb{E}[\sum_i x_i \mathbb{1}\left(Y_i = c_1\right) \mid X = x, \Theta = \theta^{(0)}]$$

$$= 0 \cdot 0.0769231 + 2 \cdot 0.25 + 2 \cdot 0.25 + 3 \cdot 0.4 = 2.2$$

$$\mathbb{E}[\sum_i x_i \mathbb{1}\left(Y_i = c_2\right) \mid X = x, \Theta = \theta^{(0)}]$$

$$= 0 \cdot 0.9230769 + 2 \cdot 0.75 + 2 \cdot 0.75 + 3 \cdot 0.6 = 4.8$$

Geometric EM

M-step For the MLE of the categorical RV Y, we normalize the expected fractional counts computed in the E-step:

$$P(Y = c_1 \mid \Theta = \theta^{(1)}) = \frac{0.9769231}{0.9769231 + 3.0230769} = 0.2442308$$

$$P(Y = c_2 \mid \Theta = \theta^{(1)}) = 1 - 0.2442308 = 0.7557692$$

We use that the MLE of this version of the geometric distribution is given by $\theta_{MLE} = \frac{n}{n+\sum_i x_i}$, but now use the fractional counts and expected sufficient statistic computed in the E-step:

new parameter of c_1 : $\dfrac{0.9769231}{0.9769231 + 2.2} = 0.3075061$

new parameter of c_2 : $\dfrac{3.0230769}{3.0230769 + 4.8} = 0.3864307$

Geometric EM

The new log-likelihood of the data $x_1, x_2, x_3, x_4 = 0, 2, 2, 3$ is computed as in a) but now using the updated parameters $\theta^{(1)} = (\theta_{c_1}^{(1)}, \theta_{c_2}^{(1)}, w_1^{(1)}, w_2^{(1)}) = (0.3075061, 0.3864307, 0.2442308, 0.7557692)$:

$$
\begin{aligned}
&\log(0.2442308 \cdot 0.3075061 \cdot 0.6924939^0 \\
&\qquad + 0.7557692 \cdot 0.3864307 \cdot 0.6135693^0) \\
&+2\cdot \log(0.2442308 \cdot 0.3075061 \cdot 0.6924939^2 \\
&\qquad + 0.7557692 \cdot 0.3864307 \cdot 0.6135693^2) \\
&+ \log(0.2442308 \cdot 0.3075061 \cdot 0.6924939^3 \\
&\qquad + 0.7557692 \cdot 0.3864307 \cdot 0.6135693^3) = -7.2323861
\end{aligned}
$$