MCSC 201

IRFAN SHEIKH

16

1.) Given dataset, $X = \{x^t, b^t\}$, $t = 1$ to $N = 7$ where x^t is 1-dimensional and $b^t = [b_1^t, b_2^t, b_3^t]$ are boolean variables denoting the cluster assignments. $b_i^t = 1$ means that instance x_i^t belongs to cluster i.

Cluster means, $m_i = \left(\sum_{t=1}^{N} x^t \cdot b_i^t\right) / \sum_{t=1}^{N} b_i^t$

- Cluster 1 : $m_1 = \dfrac{\sum_{t=1}^{N} x^t \cdot b_1^t}{\sum_{t=1}^{N} b_1^t} = \dfrac{10\cdot1 + 5\cdot0 + 18\cdot1 + 7\cdot0 + 23\cdot0 + 14\cdot1 + 19\cdot0}{1 + 0 + 1 + 0 + 0 + 1 + 0}$

$$= \frac{10 + 18 + 14}{3} = \frac{42}{3} = 14$$

Thus, $m_1 = 14$

- Cluster 2 : $m_2 = \dfrac{\sum_{t=1}^{N} x^t \cdot b_2^t}{\sum_{t=1}^{N} b_2^t} = \dfrac{5 + 7}{2} = 6$

- Cluster 3 : $m_3 = \dfrac{\sum_{t=1}^{N} x^t \cdot b_3^t}{\sum_{t=1}^{N} b_3^t} = \dfrac{23 + 19}{2} = \dfrac{42}{2} = 21$

Reconstruction Error,

$$E = \sum_{t=1}^{N} \sum_{i=1}^{3} b_i^t \, \|x^t - m_i\|^2$$

$$= 1\cdot(10-14)^2 + 1\cdot(5-6)^2 + 1\cdot(18-14)^2 + 1\cdot(7-6)^2 + 1\cdot(23-21)^2 + 1\cdot(14-14)^2 + 1\cdot(19-21)^2$$

$$= 16 + 1 + 16 + 1 + 4 + 0 + 4$$

$$= 42$$

Thus, the reconstruction error is $E = 42$.

$G_i \sim N(\mu_i\ \sigma_i)$

2.) Univariate dataset **X** : four normal distributions.

Group 1, $G_1 \sim N1(4, 18)$ — Prior probabilities: $P(G1) = 0.2$,

$G_2 \sim N2(8, 25)$ — $P(G2) = 0.3$,

$G_3 \sim N3(21, 25)$ — $P(G3) = 0.4$,

$G_4 \sim N4(31, 14)$ — $P(G4) = 0.1$

Let ϕ denote all the sufficient statistics of the distributions, and the priors. Then, for instance x, the probability $p(x)$, given the mixture model is:

$$p(x|\phi) = \sum_{i=1}^{4} p(x|G_i) \cdot P(G_i)$$

$$= p(x|G_1) \cdot P(G_1) + p(x|G_2) \cdot P(G_2) + p(x|G_3) \cdot P(G_3) + p(x|G_4) \cdot P(G_4)$$

Now, $p(x|G_i) = \dfrac{1}{\sqrt{2\pi}\,\sigma_i}\, e^{-\left(\frac{x-\mu_i}{2\sigma}\right)^2}$.

$$\Rightarrow p(x|\phi) = \frac{1}{\sqrt{2\pi}\,(18)}\, e^{-\left(\frac{x-4}{36}\right)^2} \cdot (0.2)$$

$$+ \frac{1}{\sqrt{2\pi}\,(25)}\, e^{-\left(\frac{x-8}{50}\right)^2} \cdot (0.3)$$

$$+ \frac{1}{\sqrt{2\pi}\,(25)}\, e^{-\left(\frac{x-21}{50}\right)^2} \cdot (0.4)$$

$$+ \frac{1}{\sqrt{2\pi}\,(14)}\, e^{-\left(\frac{x-31}{28}\right)^2} \cdot (0.1)$$