Quiz: Chap 7 - Clustering (MCSC201)

Max marks: 20
Max time: 30 minutes
Give formal notation for full credit.

1. Consider the following data set with given cluster assignments. Calculate the cluster mean and reconstruction error. (6)

x^t	b_1^t	b_2^t	b_3^t
10	1	0	0
5	0	1	0
18	1	0	0
7	0	1	0
23	0	0	1
14	1	0	0
19	0	0	1

2. Consider the uni-variate data set X consisting of mixture of four normal distributions - N1(4, 18), N2(8, 25), N3(21, 25), N4(31, 14), with respective mixture probabilities 0.2, 0.3, 0.4 and 0.1. Write the expression for $p(x)$ for instance x. (6)

3. Given the training set $\{XZ\}$, denoting the joint distribution of two random variables of which X is seen and Z is hidden. There are N instances in the training set $\{x^t z^t\}, t = 1, \ldots, N$. The observed part of data X is a mixture of k unknown distributions with unknown parameters represented by Φ. Distribution $\mathcal{G}_i$ has prior of π_i. Hidden part of the data Z is to be estimated using expectation maximization method. Match the expression with the most appropriate statement in this context.(8)

Expression	Interpretation
A. $\log \Pi_t p(x^t \vert \Phi)$	1. Likelihood of the hidden variable z^t for instance x^t
B. $\Sigma_{i=1}^k p(x^t \vert \mathcal{G}_i) P(\mathcal{G}_i)$	2. Log likelihood function of x
C. $P(\mathcal{G}_i \vert x^t, \Phi)$	3. Probability of cluster $\mathcal{G}_i$ for instance x^t when the mixture parameters are Φ
D. $\Pi_{i=1}^k (\pi_i)^{z_i^t}$	4. Log probability of X
	5. Probability of an instance in X
	6. Posterior probability of cluster, given the instance and distribution parameters
	7. Probability of z^t given mixture parameters are Φ