Name: Avirat Belekar
CWID:10454332
Course: CS559-B Machine Learning Fundamental and Application

Problem 2 (15pt): [Latent variable model and GMM] Consider the discrete latent variable model where the latent variable $\mathbf{z}$ use 1-of-K representation. The distribution for latent variable $\mathbf{z}$ is defined as:

$$p(z_k = 1) = \pi_k$$

where $\{\pi_k\}$ satisfy: $0 \leq \pi_k \leq 1$ and $\sum_{k=1}^{K} \pi_k = 1$. Suppose the conditional distribution of observation $\mathbf{x}$ given particular value for $\mathbf{z}$ is Gaussian:

$$p(\mathbf{x}|z_k = 1) = \mathcal{N}(x|\mu_k, \Sigma_k)$$

(1) [5pt] Write down the compact form of $p(\mathbf{z})$ and $p(\mathbf{x}|\mathbf{z})$.

(2) [5pt] Show that the marginal distribution $p(\mathbf{x})$ has the following form:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)$$

(3) [5pt] If we want to find the MLE solution for parameters π_k, μ_k, Σ_k in such model, what algorithm should we use? Briefly describe its major difference compared to K-means algorithm.

(1)

Compact form of $p(z)$ is given by:

$$p(z) = \prod_{k=1}^{K} \pi_k^{z_k} \quad \ldots\ldots (1)$$

Compact form of $p(x|z)$ is given by:

$$p(x|z) = \prod_{k=1}^{K} N(x|u_k, \Sigma_k)^{z_k} \quad \ldots (1)$$

To prove Marginal Distribution has the form

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|u_k, \Sigma_k)$$

$$p(x) = \sum_z p(z)\, p(x|z)$$

Since $p(z) = \prod_{k=1}^{K} \pi_k^{z_k}$

$$p(x|z) = \prod_{k=1}^{k} N(x|u_k, \Sigma_k)$$

Therefore $p(x) = \sum_{k=1}^{k} \pi_k N(x|u_k, \Sigma_k)$

Hence proved

$$p(x) = \sum_{k=1}^{k} \pi_k N(x|u_k, \Sigma_k)$$

(2)

MLE solution to parameters π_K, u_K, Σ_K and K-Means:-

MLE for $p(x \mid \pi, u, \Sigma)$

$$= \sum_{n=1}^{N} \ln \left\{ \sum_{K=1}^{k} \pi_K N(x \mid u_K, \Sigma_K) \right\}$$

let a random variable X be defined.

So $p(K \mid X)$

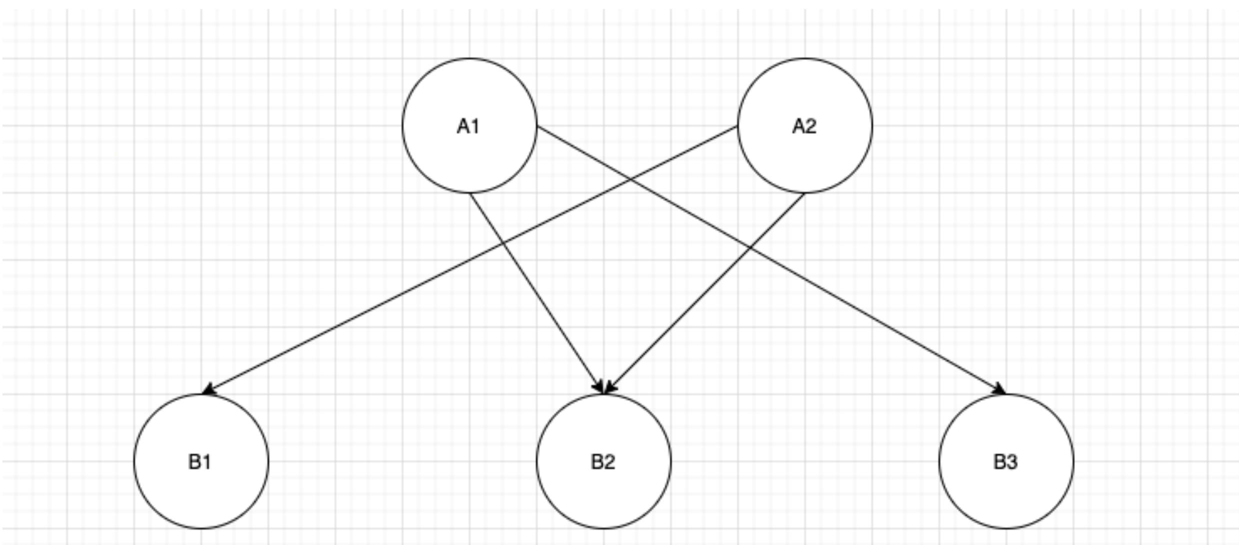
$$= \frac{p(X \mid K)\, p(K)}{\sum_{K=1}^{K} p(K)\, p(X \mid K)}$$

$$= \frac{p(X \mid K)\, \pi(K)}{\sum_{K=1}^{K} \pi_K\, p(X \mid K)}$$

(3) The Expectation-Maximization (EM) algorithm is an iterative way to find maximum-likelihood estimates for model parameters when the data is incomplete or has some missing data points or has some hidden variables. EM chooses some random values for the missing data points and estimates a new set of data. These new values are then recursively used to estimate a better first data, by filling up missing points, until the values get fixed. These are the two basic steps of the EM algorithm, namely E Step or Expectation Step or Estimation Step and M Step or Maximization Step. This is how it is different than K-Means Algorithm.

Problem 3 (20pt): [Bayesian Network] Suppose we are given 5 random variables, A1, A2, B1, B2, B3. A1 and A2 are marginally independent and B1, B2, B3 are marginally dependent on A1, A2 as follows: B1 depends on A2, B2 depends on A2 and A1. B3 depends on A1. All 5 random variables are binary, i.e., Ai, Bj ∈ {0, 1}, i = 1, 2; j = 1, 2, 3.
(1) [5pt] Draw the corresponding bayesian network.
(2) [5pt] Based on the bayesian network in (1), write down the joint distribution p (A1, A2, B1, B2, B3).
(3) [5pt] How many independent parameters are needed to fully specify the joint distribution in (2).
(4) [5pt] Suppose we do not have any independence assumption, write down one possible factorization of p (A1, A2, B1, B2, B3), and state how many independent parameters are required to describe joint distribution in this case.

Solution:
(1) Corresponding Bayesian Network



(2) Joint Distribution p (A1, A2, B1, B2, B3)
= P (A1) P (A2|A1) P(B1|A2) P (B2|B1 A2 A1) P (B3|B2 B1 A2 A1)

(3) No of independent parameter required for the joint distribution are
= P (A1) P (A2|A1) P(B1|A2) P (B2|A2 A1) P (B3| A1)
= 1 + 1 + 2 + 4 + 2 => 10
Therefore are 10 independent parameters required for the joint distribution in (2).

(4) Factorization of p (A1, A2, B1, B2, B3) is given by:
P (A1) P (A2|A1) P(B1|A2) P (B2|B1 A2 A1) P (B3|B2 B1 A2 A1)
The number of parameters required after the factorization is 1 + 2 + 4 + 8 + 16 = 31