CS 559
ASSIGNMENT 3

Name: Akshatha Vasant Hegde
CWID: 20009287

1. K means

1] What is the center of the first cluster (red) after one iteration?
[5.171, 3.171]

2] What is the center of the second cluster (green) after two iteration?
[5.3, 4.0]

3] What is the center of the third cluster (blue) when the clustering converges?
[6.2, 3.025]

4] How many iterations are required for the clusters to converge?
It took 2 iterations for the clusters to converge
I have done a third iteration to confirm the convergence.

2. Latent variable model and GMM

Consider the discrete latent variable model where the latent variable model where the latent variable z use 1-to-K representation.
The distribution for latent variable z is defined as:

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

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

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

1] Write down the compact form of p(z) and p(x|z)

→ We know that z is a binary random variable where z has 1 to k representation.
One particular element $z_k = 1$ and all other elements are 0
$\Rightarrow z_k \in \{0, 1\}$ and $\sum_k z_k = 1$

Now we know that the marginal distribution over z

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

We take an example where $K = 3$, $z = (0,0,1)$

LHS $p(z = (0,0,1)) = p(z_3 = 1) = \pi_3$

RHS $\pi_1^{z_1} \pi_2^{z_2} \pi_3^{z_3} = \pi_1^0 \pi_2^0 \pi_3^0 = \pi_3$

∴ The compact form is $p(z) = \prod_{k=1}^{K} \pi_k^{z_k}$

The conditional distribution of x given a particular value for z is a Gaussian

$$p(x \mid z_k = 1) = N(x \mid \mu_k \Sigma_k)$$

we take an example where $z = (1, 0, 0 \ldots 0)$

LHS $p(x \mid z) = p(x \mid z = (1, 0, \ldots 0)) = p(x \mid z_1 = 1) = N(x \mid \mu_1 \Sigma_1)$

RHS $N(x \mid \mu_1 \Sigma_1)^{z_1} \ldots N(x \mid \mu_k \Sigma_k)^{z_k} = N(x \mid \mu_1 \Sigma_1)^1 \ldots N(x \mid \mu_k \Sigma_k)^0$
$= N(x \mid \mu_1 \Sigma_1)^1$

∴ The compact form is $p(x \mid z) = \prod_{k=1}^{K} N(x \mid \mu_k \Sigma_k)^{z_k}$

2] Show that the marginal distribution $p(x)$ has the following form
$p(x) = \Sigma_{k=1}^{K} \pi_k N(x | \mu_k \Sigma_k)$

→ We know that $p(x) = \sum_z p(z) p(x|z)$
↳ we use Σ because z is binary and discrete

from [1] we know that
$p(z) = \pi_k^{z_k}$
$p(x|z) = N(x | \mu_k \Sigma_k)^{z_k}$
Substituting in equation we get
$p(x) = \Sigma_{k=1}^{K} \pi_k N(x | \mu_k \Sigma_k)$

3] What algorithm do we use to find MLE for parameters π_k, μ_k, Σ_k?
Breifly discribe its major difference compared to K-means algorithm?

→ With a Gaussian Density $N(x | \mu_k \Sigma_k)$ we have mean of μ_k and covariance of Σ_k
The mixing coefficiant is π_k which is, $0 \leq \pi_k \leq 1$, $\Sigma_{k=1}^{K} \pi_k = 1$
For Gaussian Mixture Model Distribution and the Marginal Distribution of GMM for discrete latent variable
$p(x) = \Sigma_{k=1}^{K} \pi_k N(x | \mu_k \Sigma_k)$
the parameters here are π, μ, Σ
The best way to find the MLE is through Expectation Management algorithm or EM algorithm.

Kmeans $\mu_k = \dfrac{\Sigma_{n=1}^{N} r_{nk} X_n}{\Sigma_{n=1}^{N} r_{nk}}$ ⇒ we see that the binary indicator (r_{nk}) is either 0 or 1

GMM $\mu_k = \dfrac{\Sigma_{n=1}^{N} r_k(X_n) X_n}{\Sigma_{n=1}^{N} r_k(X_n)}$ ⇒ Since each point has a different weight we see the weighted average $(r_{nk}(X_n))$

∴ Kmeans has hard assignments where each data point is associated uniquely with one cluster
EM for GMM have soft assigments based on the posterior probabilities.

3. Bayesian Network

Suppose we are given 5 random variables A_1, A_2, B_1, B_2, B_3
A_1 and A_2 are marginally independent. B_1, B_2, B_3 are marginally dependent on A_1 and A_2 as follows:

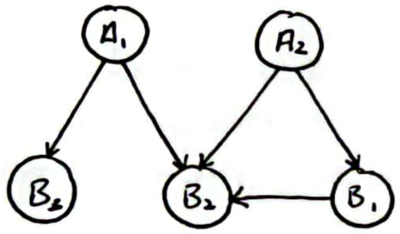
B_1 depends on A_2

B_2 depends on A_2, A_1, B_1

B_3 depends on A_1

All 5 random variables are binary ie. $A_i, B_j \in \{0, 1\}$, $i = 1,2$ $j = 1,2,3$

1] Draw the corresponding bayesian network



2] Based on the Bayesian Network in (1) write down the joint distribution

A graph with k nodes, the joint distribution is given by:
$$p(x) = \prod_{k=1}^{K} p(x_k | pa_k)$$
where pa_k denotes the parents of x_k and $x = \{x_1 \dots x_k\}$

The joint distribution of $P(A_1, A_2, B_1, B_2, B_3)$

$= P(A_1|\emptyset) P(A_2|\emptyset) P(B_1|A_2) P(B_2|A_1, A_2, B_1) P(B_3|A_1)$

$= P(A_1) P(A_2) P(B_1|A_2) P(B_2|A_1 A_2 B_1) P(B_3|A_1)$

We now factorize

$= 2^0 + 2^0 + 2^1 + 2^3 + 2^1$

$= 1 + 1 + 2 + 8 + 2$

$= 14$ parameters needed

3] How many independent parameters are needed to fully specify the joint distribution in (2)

wkt the joint distribution is

$: P(A_1)\, P(A_2)\, P(B_1|A_2)\, P(B_2|A_1, A_2, B_1)\, P(B_3|A_1)$

Here we know that all probabilities are conditionally independent
We have 5 conditionally independent probabilities.

where $P(A_1)\, P(A_2)\, P(B_1|A_2)\, P(B_2|A_1, A_2, B_1)\, P(B_3|A_1)$

$= 2^0 + 2^0 + 2^1 + 2^3 + 2^1$

$= 1 + 1 + 2 + 8 + 2$

$=$ 14 parameters needed

4] Suppose we do not have any independent assumption, write down one possible factorization of $p(A_1, A_2, B_1, B_2, B_3)$ and state how many independent parameters are required to describe joint distribution in this case.

For any set of random variables, the probability of any member of a joint distribution can be calculated from the conditional probabilities using the chain rule

The chain rule/ product rule is:

let $S_1, S_2 \ldots S_n$ be events and $p(S_1) > 0$ then:

$p(S_1, S_2 \ldots S_n) = P(S_1)\, P(S_2|S_1) \ldots p(S_n|S_{n-1}, \ldots S_1)$

For $P(A_1, A_2, B_1, B_2, B_3)$, applying the chain rule, the joint distribution :

$= P(A_1)\, P(A_2|A_1)\, P(B_1|A_2, A_1)\, P(B_2|B_1, A_2, A_1)\, P(B_3|B_2, B_1, A_2, A_1)$

Factorizing :

$= 2^0 + 2^1 + 2^2 + 2^3 + 2^4$

$= 1 + 2 + 4 + 8 + 16$

$=$ 31 parameters.

$\therefore 2^n - 1 = 2^5 - 1 = 32 - 1 = 31$

4] Neural Network

In code