CSE 575: Homework #3

Amey Bhilegaonkar

11/04/2022

Problem 1

a)

Solution: Every variable V is intended to be conditionally independent of Non-descendants(V) given parents, according to the Markovian assumption (V).
Hence, in our Bayesian Network:

Node	Parents	Non-Descendants
A	{}	{}
B	{ A }	{C}
C	{A}	{B}
D	{B,C}	{A,B,C}

Therefore, by using Markovian Assumption:

- A is independent of No variables
- B is independent of C given A.
- C is independent of B given A.
- D is independent of A given B,C.

b)

Solution: A node's x_i set of parents, children, and co-parents make up the node's Markov blanket. It has the property that only the variables in the Markov blanket affect the conditional distribution of x_i, which is dependent on all the other variables in the graph.
Hence, in our Bayesian Network:

Node	Markov's Blanket
A	{B,C}
B	{ A,C,D }
C	{A,B,D}
D	{B,C}

Therefore, by using Markov's Blanket on each variable we can summarize:

- A is independent of D given B,C.
- B is independent of no variables.

- C is independent of no variables.
- D is independent of A given B,C.

c)

Solution: We know that joint distribution over K variables given by $P(x_1, x_2, ..., x_K)$ can be written as a product of conditional distributions, one for each of the variables. The joint distribution can be represented as follows:
$P(x_1, x_2, ..., x_K) = P(x_k|x_1, x_2, ..., x_K - 1)\ ...\ P(x_2|x_1)\, P(x_1)$
The factorization of the joint probability distribution over A, B, C, and D can be expressed as follows using the equation mentioned above for the given problem:
$P(A, B, C, D) = P(D|B, C)\, P(B|A)\, P(C|A)\, P(A)$

d)

Solution: Yes, the joint probability table given to us is this a valid distribution for the Bayesian network.
Reason: The probability distribution of D discrete variables, namely X_1, X_2, ..., X_D, each having m values. For the distribution to be valid for a Bayesian Network, the joint probabilities for all possible combinations of values of X_1, X_2, ..., X_D should sum to 1.

$\sum_{X_1, X_2, ..., X_D} P(X_1, X_2, ..., X_D) = \prod_{i=1}^{D} \sum_{X_i} P(x_i|Parents(x_i))$

from the table above in the question: $\sum_{X_i} P(x_i|Parents(x_i)) = 1$

Therefore, it would be a valid distribution for the Bayesian Network if the total of the probabilities provided to us in the Joint Probability Distribution Table (Table 1) adds up to 1.

Sum of all values is 1

Problem 2

a)

Solution: $P(A = 1|B = 1, C = 0) = \frac{P(B=1,C=0|A=1)*P(A=1)}{P(B=1,C=0)}$
$= \frac{P(B=1|A=1)*P(C=0|A=1)*P(A=1)}{(P(B=1,C=0|A=1)*P(A=1)+P(B=1,C=0|A=0)*P(A=0))}$
$= \frac{\theta_{B|1}*(1-\theta_{C|1})*\theta_A}{(\theta_{B|1}*(1-\theta_{C|1})*\theta_A+\theta_{B|0}*(1-\theta_{C|0})*(1-\theta_A))}$

b)

Solution: Likelihood can be given as:

$P(A, B, C, D) = \pi_i P(D_i|B_i, C_i)P(B_i|A_i)P(C_i|A_i)P(A_i)$

$\ln(P(A, B, C, D)) = \sum(\ln P(D_i|B_i, C_i) + \ln P(B_i|A_i) + \ln P(C_i|A_i)\ +\ \ln\ P(A_i))$

$Loglikelihood(\theta|D) = 3ln(\theta_A) + 5ln(1 - \theta_A) + 3ln(\theta_{B|A=0}) + 2ln(1 - \theta_{B|A=0}) + 2ln(\theta_{B|A=1}) + ln(1 - \theta_{B|A=1}) + 3ln(\theta_{C|A=0}) + 2ln(1 - \theta_{C|A} + ln(\theta_{C|A=1}) + 2ln(1 - \theta_{C|A=1}) + 2ln(1 - \theta_{D|B=0|C=0}) + 1ln(\theta_{D|B=0|C=1}) + 2ln(\theta_{D|B=1|C=0}) + 2ln(\theta_{D|B=1|C=1}) + 1ln(1 - \theta_{D|B=1|C=1})$

c)

Solution:
1) Differentiating and equating to zero we get,
$\frac{\partial LogL(\theta|D)}{\partial\theta_A} = \frac{3}{\theta_A} - \frac{5}{(1-\theta_A)} = 0$

$\theta_A = 3/8$

2) $\frac{\partial LogL(\theta|D)}{\partial\theta_{B|A=0}} = \frac{3}{\theta_{B|A=0} - 2/(1-\theta_{B|A=0})} = 0$

$\theta_{B|A=0} = 3/5$

3) $\frac{\partial LogL(\theta|D)}{\partial\theta_{B|A=1}} = \frac{2}{\theta_{B|A=1} - 1/(1-\theta_{B|A=1})} = 0$

$\theta_{B|A=0} = 2/3$

4) $\frac{\partial LogL(\theta|D)}{\partial\theta_{C|A=0}} = \frac{3}{\theta_{C|A=0}} - \frac{2}{(1-\theta_{C|A=0})} = 0$
$\theta_{C|A=0} = 3/5$

5) $\frac{\partial LogL(\theta|D)}{\partial\theta_{C|A=1}} = \frac{1}{\theta_{C|A=1}} - \frac{2}{(1-\theta_{C|A=1})} = 0$
$\theta_{C|A=0} = 1/3$

6) $\frac{\partial LogL(\theta|D)}{\partial\theta_{D|B=0,C=0}} = \frac{0}{\theta_{D|B=0,C=0}} - \frac{2}{(1-\theta_{D|B=0,C=0})} = 0$
$\theta_{D|B=0,C=0} = 0$

7) $\frac{\partial LogL(\theta|D)}{\partial\theta_{D|B=0,C=1}} = \frac{1}{\theta_{D|B=0,C=1}} - \frac{0}{(1-\theta_{D|B=0,C=1})} = 0$
$\theta_{D|B=0,C=1} = 1$

8) $\frac{\partial LogL(\theta|D)}{\partial\theta_{D|B=1,C=0}} = \frac{2}{\theta_{D|B=1,C=0} - \frac{0}{(1-\theta_{D|B=1,C=0})}} = 0$
$\theta_{D|B=1,C=0} = 1$

9) $\frac{\partial LogL(\theta|D)}{\partial\theta_{D|B=1,C=1}} = \frac{2}{\theta_{D|B=1,C=1}} - \frac{1}{(1-\theta_{D|B=1,C=1})} = 0$
$\theta_{D|B=1,C=1} = \frac{2}{3}$

d)

Solution: For the Bayesian Network in Figure 1 the Likelihood is given by :
$P(A,B,C,D) = P(D|B,C)\,P(C|A)\,P(B|A)\,P(A)$

For the Bayesian Network in Figure 2 the Likelihood is given by:
$P(A,B,C,D) = P(D|B,C,A)\,P(C|A,B)\,P(B|A)\,P(A)$

Adding of more dependency between variables changes the likelihood for C and D while it remains the same for A and B.

By observation, $P(D|B, C)$ is actually $P(D|A, B, C)$ when the variable A is marginalized similarly $P(C|A)$ is actually $P(C|A, B)$ marginalized over B

Thus the Alternate Bayesian Network Likelihood is given in Figure 2 is **at least** that of the Bayesian Network in Figure 1.

Problem 3

a)

Solution: $p(X) = \Sigma_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k)$

$|\Sigma_k| = \sigma_{k1}^2 + \sigma_{k2}^2 ... + \sigma_{kD}^2$

$N(x|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{\frac{1}{2}} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)$

$= 1/(2\pi)^{D/2} |\Sigma_k|^{-} 0.5 e^{(-0.5 \Sigma_i (x_i - \mu_{ki}/\sigma_{ki}^2)(x - \mu_k))}$

$= 1/(2\pi)^{D/2} |\Sigma_k|^{-} 0.5 \Pi_{i=1}^{D} \exp{(x_i - \mu_{ki}/\sigma_{ki})^2}$

$= \Pi_{i=1}^{D} (1/\sigma_{ki} * \sqrt{2\pi}) \exp{(x_i - \mu_{ki}/\sigma_{ki})^2}$

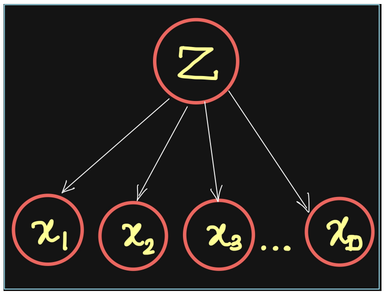
$= \Pi_{i=1}^{D} N(x_i|\mu_{ki}, \sigma_{ki}^2)$

Hence $f_{ki}(x)$ = $N(x_i|\mu_{ki}, \sigma_{ki}^2)$

b)

Solution: Since, we know that the GMM's covariance matrix is diagonal matrix it implies that all the random variables are independent.
Thus the Bayesian Network for the Simplified GMM can be represented as follows :

Problem 3 continued on next page. . .



The conditional probabilities can be given as:
$P\,(Z = k) = \pi_k$
$P(x_i|z = k) = N(x_i|\mu_{ki}, \sigma^2_{ki})$ where $i\ \in\ [1, D]$

c)

Solution:
P(z_n = $k|x_n$) = $\pi_k N(x_n|\mu_k, \sum_k)/(\sum_{k-1} \pi_k N(x_n|\mu_k, \sum_k)$
=$\pi_k \prod_{i-1}^{D} N(x_i|\mu_{ki}, \sigma^2_{ki})/(\sum_{k-1} \pi_k \prod_{i-1}^{D} N(x_i|\mu_{ki}, \sigma^2_{ki}))$
M Step:
$\Sigma_k = \frac{\Sigma_{n=1}^{N} \gamma(z_n=k).(x_n-\mu_k).(x_n-\mu_k)^T}{\Sigma_{n=1}^{N} \gamma(z_n=k)}$
$\sigma^{2new}_{ki} = \frac{\Sigma_{n=1}^{N} \gamma(z_n=k).(x_{ni}-\mu_{ki})^2}{\Sigma_{n=1}^{N} \gamma(z_n=k)}$

d)

Solution: The parameter values after one Expected Maximization are as follows:
$\pi_1 = 0.4317$
$\pi_2 = 0.568$
$\mu_{11} = 1.414$
$\mu_{12} = 0.2387$
$\mu_{21} = -0.2625$
$\mu_{22} = -0.04062$
$\sigma^2_{11} = 0.77$
$\sigma^2_{12} = 3.15$
$\sigma^2_{21} = 2.38$
$\sigma^2_{22}\ =\ 5.04$