CSE 575: Homework #3

Due: November 4, 2022

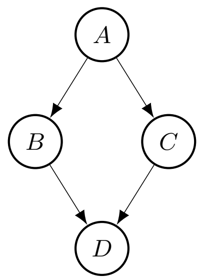

Figure 1: Bayesian network

Problem 1

Consider the Bayesian network structure given in Figure 1.

a) (6pt) Use the Markovian assumption on each variable to derive independence statements described by this Bayesian network.

b) (6pt) For each variable, **(i)** identify its Markov blanket and **(ii)** state any independence that can be derived using the Markov blanket.

c) (6pt) Write down the factorization of the joint probability distribution over A, B, C, D that corresponds to this structure.

d) (6pt) Consider the joint probability table in Table 1. Is this a valid distribution for the Bayesian network in Figure 1? Explain your answer.

Problem 2

Again consider the Bayesian network structure in Figure 1 and the associated conditional probability tables (CPTs) below. Assume all variables are binary.

$P(A=1)$
θ_A

A	$P(B=1\|A)$
0	$\theta_{B\|0}$
1	$\theta_{B\|1}$

A	$P(C=1\|A)$
0	$\theta_{C\|0}$
1	$\theta_{C\|1}$

B	C	$P(D=1\|B,C)$
0	0	$\theta_{D\|0,0}$
0	1	$\theta_{D\|0,1}$
1	0	$\theta_{D\|1,0}$
1	1	$\theta_{D\|1,1}$

A	B	C	D	$P(A,B,C,D)$
0	0	0	0	1/16
0	0	0	1	1/16
0	0	1	0	1/16
0	0	1	1	1/16
0	1	0	0	1/32
0	1	0	1	1/32
0	1	1	0	3/32
0	1	1	1	3/32
1	0	0	0	5/64
1	0	0	1	7/64
1	0	1	0	5/64
1	0	1	1	7/64
1	1	0	0	1/32
1	1	0	1	1/32
1	1	1	0	1/32
1	1	1	1	1/32

Table 1: Joint distribution table

A	B	C	D
0	1	1	0
1	1	1	1
1	0	0	0
0	1	1	1
1	1	0	1
0	0	0	0
0	1	0	1
0	0	1	1

Table 2: Training data

a) (6pt) Write down the expression for the probability $P(A = 1 \mid B = 1, C = 0)$ in terms of the parameters shown in the above CPTs.

b) (10pt) Suppose you are given the training data in Table 2. What is the log-likelihood function of above Bayesian network given this data? Note: it should be a function of the parameters shown in the above CPTs.

c) (10pt) Optimize the log-likelihood function from part (b) to obtain the maximum-likelihood parameters for this Bayesian network. Please show your work.

d) (6pt) Consider an alternative Bayesian network structure shown in Figure 2. Would the likelihood achieved by this BN be at least that of the BN from part (c), or would it be smaller? Justify your answer.

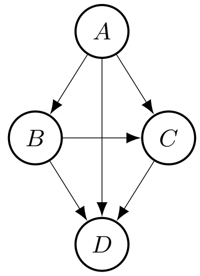

Figure 2: Bayesian network

Problem 3

Recall that a Gaussian mixture model (GMM) defines a probability density function as follows:

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

where π_k are the mixing coefficients satisfying $\pi_k \geq 0, \sum_{k=1}^{K} \pi_k = 1$, and $\mathcal{N}(\cdot \mid \mu_k, \mathbf{\Sigma}_k)$ denotes a multivariate Gaussian with mean μ_k and covariance $\mathbf{\Sigma}_k$.

Consider a simplified GMM whose covariance matrices are diagonal. In other words, for $k \in \{1, \ldots, K\}$, the covariance matrix $\mathbf{\Sigma}_k$ can be expressed as:

$$\mathbf{\Sigma}_k = \begin{bmatrix} \sigma_{k1}^2 & & & \\ & \sigma_{k2}^2 & & \\ & & \ddots & \\ & & & \sigma_{kD}^2 \end{bmatrix}$$

where D is the dimension of $\mathbf{x}$. Moreover, denote each entry of μ_k as μ_{ki} for $i \in \{1, \ldots, D\}$. That is, $\mu_k = (\mu_{k1}, \ldots, \mu_{kD})^T$.

a) (10pt) Show that the density function of such simplified GMM can be expressed in the following form:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \prod_{i=1}^{D} f_{ki}(\mathbf{x}).$$

Derive an expression for $f_{ki}(\mathbf{x})$ in terms of $\mathbf{x}, \mu_{ki}, \sigma_{ki}^2$.

b) (10pt) Recall the latent variable interpretation of GMMs, by introducing a discrete latent variable z that can take values in $1, \ldots, K$. **(i)** What would the Bayesian network structure of this simplified GMM be? The DAG should contain nodes corresponding to $z, x_1, \ldots, x_D$. **(ii)** Describe the conditional probability table/densities of this Bayesian network using $\pi_k, \mu_{ki}, \sigma_{ki}^2$.

c) (12pt) You will now derive the Expectation Maximization (EM) algorithm for this simplified GMM given a set of data examples $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$. **(i)** Derive the expression for $\gamma(z_n = k)$ (i.e., $p(z_n = k \mid \mathbf{x}_n)$) for the E step. **(ii)** Derive the update rule for σ_{ki}^2 in the M step. Your answer may rely on the value of μ_{ki} in the current M step.

d) (12pt) Consider the training data shown in Table 3. Assume that $K = 2$ and $D = 2$. Suppose the

Problem 3 continued on next page...

x_1	x_2
1.1	0.4
-2	1
0	0.5
1	-1.5
2.2	0

Table 3: Training data

parameters are initialized as follows:

$$\pi_1 = 0.2$$
$$\mu_{ki} = \begin{cases} i & \text{if } k = 1 \\ -i & \text{if } k = 2 \end{cases}$$
$$\sigma_{ki}^2 = 0.4 \times k \times i$$

Show the parameter values after the one iteration of EM.