**Birla Institute of Technology & Science, Pilani**

**Work Integrated Learning Programmes Division**

**First Semester 2022-2023**

**Comprehensive Examination**

**(EC-3 Makeup)**

Course No. : DSECLZG526

Course Title : Probabilistic Graphical Models

Nature of Exam : Open Book

Weightage : 40% (As per Course Handout)

No. of Pages = 3

No. of Questions = 6

# No. of Questions = 6

Duration : 2.5 Hours

Date of Exam : 29/04/2023 (AN)

Note to Students:

1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.
3. Assumptions made if any, should be stated clearly at the beginning of your answer.

(1a) A pair of data analysts working at a company built a Chow-Liu tree on data consisting of a set of variables with considerable effort. The customer for which the tree is being constructed introduces a new variable after the tree is constructed and wants the new data to be incorporated in the tree. The first data analyst looks at the mutual information measures between the new variable and all the existing variables in the data and finds that for the new variable if and then The first data analyst then claims that the new Chow-Liu tree can be constructed by making a small modification to the old Chow-Liu tree, while the other analyst disagrees and says that the new Chow-Liu tree must be constructed from scratch and will be quite different from the old Chow-Liu tree. Who is right? Give detailed justification for your answer.

[5 Marks]

(1b) The first data analyst goes on to claim that if the values of and are defined as follows: and and are found to obey the relationship , we would not need to run the Chow-Liu algorithm to construct the new Chow-Liu tree. Is the first data analyst right in this case as well? Give detailed justification for your answer.

[5 Marks]

Solution

1. The given information of mutual information values ensures that when the mutual information pair values are sorted in descending order, the new mutual information values of the form will be below those of the original mutual information pair values of the form . Kruskal’s algorithm for the maximum spanning tree will then construct the old spanning tree since it sees all the old edges first and then pick an edge where is a variable such that . Thus the first data analyst is right in saying that the new tree is not significantly different from the old tree, it is just the old tree + one extra edge.

Marking Scheme: 3 Marks 🡪 for showing how Kruskal’s algorithm would work in this case, 2 Marks 🡪 what does the final MST look like, and which analyst is correct.

1. In this case the given mutual information pair values are such that when they are sorted in descending order all mutual information pairs of the form will appear first followed by old mutual information values such as Applying Kruskal’s algorithm in this case would give rise to a star-shaped tree centered on and with edges So the first analyst is right in saying that there is no need to run the Chow-Liu algorithm again – the tree can be seen to be a star-shaped tree without construction.

Marking Scheme: 3 Marks 🡪 for showing how Kruskal’s algorithm would work in this case, 2 Marks 🡪 what does the final MST look like, and which analyst is correct.

1. Let be a Bayesian network on the set of variables . Let be the set of leaf nodes in . Removing the set of nodes from we obtain the network . Show that the probability distribution of in is the same as the probability distribution on those variables in , i.e Give detailed steps in your derivation. [6 Marks]

Solution

Let be a leaf node. Then the probability distribution over the Bayesian network on can be written as Note that .

Thus the distribution in on is the same as the distribution on those variables in which is obtained from by removing the leaf node . Now we can remove a leaf node from to get while preserving the distribution on the remaining variables.

Thus we can remove all nodes in from to get a network such that

Marking Scheme: Main derivation showing that removing one leaf node preserves the distribution on the remaining variables in the residual network 🡪 5 marks, remaining argument 🡪 1 mark

1. Apply the max-product variable-elimination algorithm on the graph below using the elimination orderings and . List all the intermediate steps in the computation. Assume that each of the variables is binary-valued, and find out the size of the largest factor formed in each case. [6 Marks]

B

A

E

D

C

H

G

F

Solution

The Bayesian factorization for the joint probability distribution of the given network can be written as

Let us consider the elimination order first.

Eliminating : and .

Eliminating :

Eliminating :

Eliminating : ,

Eliminating :

Eliminating :

Eliminating :

Eliminating :

The largest factor formed is the one on the variables which has rows in it.

Let us now consider the elimination ordering .

Eliminating : and .

Eliminating :

Eliminating :

Eliminating :

Eliminating :

Eliminating :

Eliminating :

Eliminating :

The largest factor formed in this case is on 3 variables, which means that the largest factor has rows.

Suggested Marking Scheme

First elimination ordering 🡪 3 Marks

Second elimination ordering 🡪 3 Marks

1. Suppose there is a wallet containing coins containing five one-rupee coins, five 5-rupee coins, and five 10-rupee coins. Coins are drawn one at a time randomly from the wallet and placed on a table. Let represent the total value of the coins on the table after draws. Let Can the sequence be modeled as a Markov chain? Give detailed justification for your answer.

[6 Marks]

Solution

We need to check if knowledge of the value of is sufficient to tell what coins were used in these draws. If that is so, then we have a Markov chain, because the probabilities of depend only on and we know which coins are left to be drawn. On the other hand, if knowledge of the value of alone does not give us only one way of getting that value, then we will need to know the previous states also in order to calculate the probability of since we will have multiple possibilities for coins that are left to be drawn.

Let us assume that there are two ways of drawing the value in draws. Thus we seek two triples and such that each entry in the tuple is an integer between 0 and 5, and stands for the number of coins of the respective kind that are selected. We have and

Rearranging the first equation we have and rearranging the second equation we Now subtracting these two equations we get . From here we conclude that and where is an integer.

After some more manipulation we arrive at . Note that we have For a valid we see that which means that since When we see that , so this rules out When we see that since The only option for is Thus and there is only one solution to the equations which means there is only way to get the value represented by which means it is a Markov chain.

Marking Scheme: 1 Mark 🡪 stating that the sequence is a Markov chain, 5 Marks 🡪 explanation for why it is a Markov chain.

1. Consider the Bayesian network below. Suppose we marginalize over the variable . Construct a Bayesian network to represent a minimal I-map for the marginal distribution of the remaining variables defined by the network. Give suitable justification for your construction.

[6 Marks]

D

E

A

X

B

G

F

C

H

Answer

Marginalizing over variable means not observing it. Thus it means we need to remove the node and all edges incident upon it, and add new edges to preserve dependency relationships among the variables.



First we add edges between the parents of and the children of , since is unobserved and creates an active path between these nodes. We also add edges between the children of . The direction of edges from parents of to children of is from the parent of to the child of , but the direction of edges between children of is arbitrary. Thus we add the edges and



Two additional edges needs to be added - in the problem graph above specifying creates an active path between and and between and since is a converging node and is unobserved. However in the graph constructed so far without the edge from to , and from to these active paths do not exist. We therefore add an edge between and , and between and

Note that this is not the only graph that can be constructed – we can also construct a graph very similar to the above with the direction on the edge reversed, and a new edge added from to in place of one from to

Suggested Marking Scheme:

4 Marks 🡪 adding edges between parents of and children of and between children of

2 Marks 🡪 recognizing the need for the edge from to , and from to

1. The best prior information available with us about a coin with sample-space suggests that its parameter vector obeys a Dirichlet distribution with hyperparameters 7 and 1 respectively. We observe a certain data consisting of an unknown number of tosses of the coin and find that the posterior probability is What is the smallest value can take that would explain this observation? For this value of , what is the largest value of that we would be able to see for any data consisting of tosses?

[6 Marks]

Solution

We have where

Thus we have

After some manipulation we arrive at the equation . Since , we have . takes its smallest value when takes its smallest value, which is 0.

Thus The largest value for for any data will occur when consists of all heads, i.e when in This means that the largest value for