UNIVERSITY OF LINCOLN  SCHOOL OF COMPUTER SCIENCE

CMP9132M ASSESSMENT ITEM 1

ADVANCED ARTIFICIAL INTELLIGENCE

NAME = SADANAND VITHAL GIRADDI

STUDENT ID = 25895786

Extension Authorization code -

KZLWPX38SFYFDRXC

Table of Contents

ADVANCE ARTIFICIAL INTELLIGENCE

1. PROBLEM DEFINITION

Dataset of patients has been provided with the problem, the first part of the problem is to obtain the Causal Bayesian Network (CBN) and its conditional probability tables (CPTs) using the given data. The second part of the problem is to implement the python program that automatically computes the intervention's probability of P (Y | do X) for all the values of X and Y. The final part of the problem is to introduce another variable which is used to check the patient mental problem by creating the new CBN that computes the Average causal effect (ACE) with the given different data.

2. OBJECTIVE

The main objective of this problem is to understand the concept of Bayesian network and implementing the dataset in the python program by finding the probability of all the possible values by adjusting the formulas for the intervention and also finding the status of the patient mental problem using Average causal effect for the given problem.

3. PROCEDURE

3.1. First Part Solution

In first part of the problem, dataset of the patients has been provided in which the treatment X (i.e placebo or medicine) depends upon the Z (Mental health), Y (Survival cases) depends both on the treatment X and Z (Mental health), Z (Mental health) acts as a parent. The graphical representation of the Causal Bayesian Network (CBN) and its conditional probability tables (CPTs) becomes as mentioned in result.

For conditional probability tables (CPTs) the values are imported from the given dataset csv file and found all the possible values for X, Y, Z using the python program. Libraries used are – pandas, numpy to make mathematical operations.

df.groupby('X').count() is used to find the count of occurrences in the X using the groupby() Method. And same method is used to find the counts for the Y and Z values.

And after finding the counts of Occurrences in the X, Y, and Z. The values are later divided by 1000 because the dataset has 1000 rows inorder to find the probability of each Occurrences.

Treatment = X [X=0 for nothing, X=1 for placebo, X=2 for medicine]

Survival cases = Y [Y=0 for ¬survive and Y=1 for survive]

Mental Health = Z [Z=0 for ¬healthy and Z=1 for healthy]

Now, the probability of each values is pX0= 0.171 pX1= 0.436 pX2= 0.393 pY0= 0.254 pY1= 0.746 pZ0= 0.272 pZ1= 0.728

Similarly, To create the conditional probability table for Causal Bayesian Network(CBN) the formula is:

```
pofX0Z1=(len(df[(df.X==0) & (df.Z==1)]))/1000
pX0Z1=pofX0Z1/pZ1
print(f"the value of P(X0/Z1)={pX0Z1}")
```

Output- the value of P(X0/Z1)=0.1978021978021978

```
pofX0Z0=(len(df[(df.X==0) & (df.Z==0)]))/1000
pX0Z0=pofX0Z0/pZ0
print(f"the value of P(X0/Z0)={pX0Z0}")
```

Output- the value of P(X0/Z0)=0.09926470588235294

From the conditional probability chain rule formula:

```
pofY1X0Z1=len(df[(df.Y==1) & (df.X==0) & (df.Z==1)])/1000 pY1X0Z1= pofY1X0Z1/pofX0Z1
print(f"the value of p(Y1/X0,Z1)={pY1X0Z1}")
```

Output-the value of p(Y1/X0,Z1)=0.80555555555557

```
pofY0X0Z1=len(df[(df.Y==0) & (df.X==0) & (df.Z==1)])/1000 pY0X0Z1= pofY0X0Z1/pofX0Z1
print(f"the value of p(Y0/X0,Z1)={pY0X0Z1}")
```

Output-

the value of p(Y0/X0,Z1)=0.19444444444444448

Probability = p(Y1/X0,Z1)+ p(Y0/X0,Z1)= 1

(should be equal to the value 1).

Likewise, the values for other probabilities of each values of X, Y, Z are mentioned in the tabular column in the result for First part solution.

3.2. Second Part Solution

In the second part of the problem, the problem has to be implemented and document a python program that reads the dataset and computes it into the intervention's probability of P(Y | do(X)) for all the possible values of X and Y by importing the data from the csv file.

The causal effect P (Y | do(X)) is equal to the conditional probability Pm(Y | X) in the manipulated model, i.e. P(y | do(x)) = P m (y | x)

From reference [1] and [2] as well as from classroom presentation the below formulas are as follows for the intervention probability.

$$
\begin{aligned}
P(y|do(x)) &= P_m(y|x) && \text{from definition of intervention}\\
&= \sum_z P_m(y|x,z)P_m(z|x) && \text{from Law of Total Probability}\\
&= \sum_z P_m(y|x,z)P_m(z) && \text{from independence of } X \text{ and } Z\\
&= \sum_z P(y|x,z)P(z) && \text{from previous slide's equalities}
\end{aligned}
$$

For conditional probability tables (CPTs) the values are imported from the given dataset csv file and found all the possible values for X, Y, Z using the python program. Libraries used are – pandas, numpy to make mathematical operations.

By adjusting the formula from reference [2]

P(Y/doX)=(P(Y/X,Z))*P(Z)+(P(Y/X,-Z))*P(-Z)

Now, for finding the values for X and Y,

For X=0 Y=1,

```
pofY1doX0=(pY1X0Z1*pZ1)+(pY1X0Z0*pZ0)
print(f"the value of p(Y1/(doX0))={pofY1doX0}")
```

Output-

the value of p(Y1/(doX0))=0.7174074074074075

Similarly, For X=0 Y=0,

Output-

the value of p(Y0/(doX0))=0.2825925925925926

Probability = p(Y1/(doX0)) + p(Y0/(doX0)) = 1

(equal to the value 1).

Similarly, the values for other possible probabilities of each values of X, Y, Z are mentioned in the tabular column in the result for Second part solution.

3.3. Final Part solution

In the final part of the problem, the problem has to modify the CBN to introduce another variable W ∈ {gene, ¬gene} indicating whether the patient is genetically predisposed to some mental problem by influencing the mental condition(Z).

According to the following probabilities:

P(Z = healthy | W = gene) = 0.6 P(Z = healthy | W = ¬gene) = 0.8 P(W = gene) = 0.1

But I assumed the following values like Z [Z=1 for healthy]

W = gene notW = ¬gene

Z1= healthy Z0 = ¬healthy

Formula To find the (average causal effect)

ACE = (Y | do(W))- (Y | do(-W)) from reference [3]

PofZ1W= 0.6

PofZ1notW=0.8

PofZ0W= 1-0.6 (the law of probability states that P(Z)+P(-Z)=1)) from reference [1]

PofZ0notW=1-0.8 (the law of probability states that P(Z)+P(-Z)=1))

By the enumeration approach from the causal inference in the Bayesian network formula obtained is:

```
pofY1doW=(pX0Z1*pY1X0Z1*PofZ1W)+(pX0Z0*pY1X0Z0*PofZ0W)+(pX1Z1*pY1X1Z1*PofZ1W)+(pX1Z0*pY1X1Z0*PofZ0W)+(pX2Z1*pY1X2Z1*PofZ1W)+(pX2Z0*pY1X2Z0*PofZ0W)
print(f"the value of P(Y1/doW)={pofY1doW}")
pofY1donotW=(pX0Z1*pY1X0Z1*PofZ1notW)+(pX0Z0*pY1X0Z0*PofZ0notW)+(pX1Z1*pY1X1Z1*PofZ1notW)+(pX1Z0*pY1X1Z0*PofZ0notW)+(pX2Z1*pY1X2Z1*PofZ1notW)+(pX2Z0*pY1X2Z0*PofZ0notW)
print(f"the value of P(Y1/(do-W))={pofY1donotW}")
ACE = pofY1doW - pofY1donotW
print(f"the value of Average causal effect (ACE)={ACE}")
print(f"the patient is genetically predisposed to some mental problem because of negative effects on the patient survival i.e (ACE<0)
```

Output-

the value of P(Y1/doW)=0.7105042016806722

the value of P(Y1/(do-W))=0.765966386554622

the value of Average causal effect (ACE)

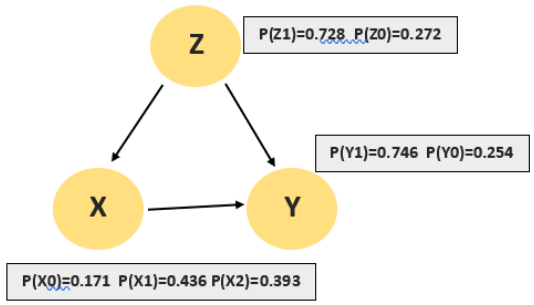
= -0.05546218487394983

the patient is genetically predisposed to some mental problem because of negative effects on the patient survival i.e (ACE<0)

4. RESULTS

The graphical representation of the Bayesian network and the conditional probability table is obtained from the given dataset and the probability of all the possible values of X, Y, Z has been found and created the conditional probability table (CPT) using the suitable formulas and with different methods.

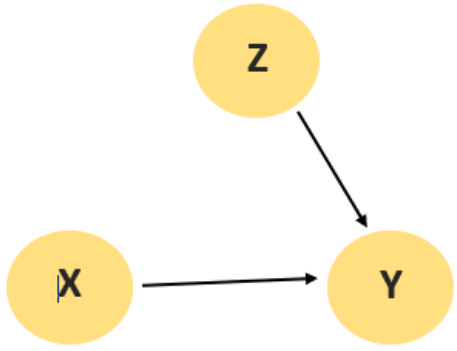
First part solution



Z	P(X0|Z1)	P(X1|Z1)	P(X2|Z1)
T=1	0.197	0.4409	0.361
	P(X0|Z0)	P(X1|Z0)	P(X2|Z0)
F=0	0.099	0.4227	0.4779

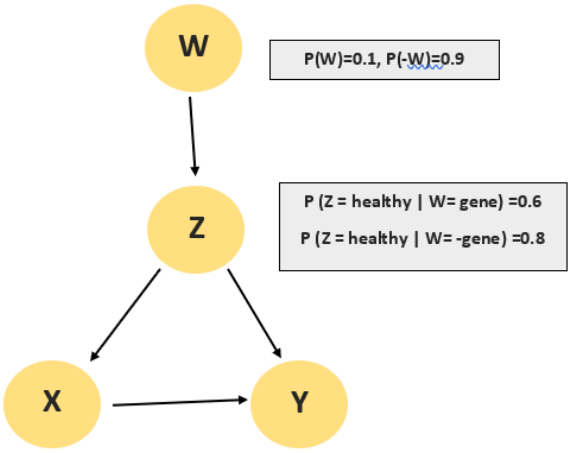
Z	P(Y1|X0,Z)	P(Y1|X1,Z)	P(Y1|X2,Z)	P(Y0|X0,Z)	P(Y0|X1,Z)	P(Y0|X2,Z)
T=1	0.8055	0.7570	0.9087	0.1944	0.2429	0.09
F=0	0.4814	0.5478	0.5538	0.5185	0.4521	0.44

Second part solution



Y	P(Y1|doX0)	P(Y1|doX1)	P(Y1|doX2)
T=1	0.717	0.700	0.812
	P(Y0|doX0)	P(Y0|doX1)	P(Y0|doX2)
F=0	0.282	0.299	0.187

Final part solution



the value of P(Y1/doW)=0.7105042016806722

the value of P(Y1/(do-W))=0.765966386554622

the value of Average causal effect (ACE)

= -0.05546218487394983

he patient is genetically predisposed to some mental problem because of negative effects on the patient survival i.e. (ACE<0)

5. CONCLUSION

The number of possible outcomes has been found by using the dataset and also shows how to use the formulas for specific problem in solving with mathematical equations. And also known about the execution of the codes, reading the data from the dataset and importing it into the python program.

6. REFERENCES

[1] Pearl et al. “Causal Inference in Statistics -- A Primer” ○ Sec 1.3.1-1.3.9 “Probability and Statistics”.

[2] Pearl, Glymour & Jewell “Causal Inference in Statistics -- A Primer” ○ Ch 1 - Sec 1.1, 1.2 ○ Ch 2 ○ Ch 3 - Sec 3.1, 3.2

[3] Pearl et al. “Causal Inference in Statistics -- A Primer” ○ Sec1.4 ○ Sec 1.5.2