Advanced Artificial Intelligence, 2425

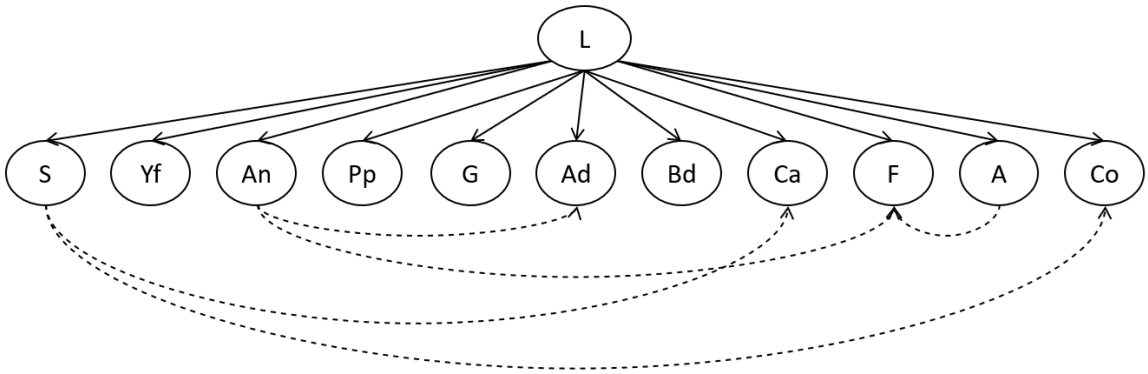
Workshop Task: Quiz on Structure Learning

Question 1

20 points

Consider the following Bayes Net structure for the Lung Cancer dataset in your data folder (the long names of random variables are the actual ones used in the dataset):

S=Smoking	Bd=Born_an_Even_Day
Yf=Yellow_Fingers	Ca=Car_Accident
An=Anxiety	F=Fatigue
Pp=Peer_Preassure	A=Allergy
G=Genetics	Co=Coughing
Ad=Attention_Disorder	L=Lung_cancer



Make use of a Conditional Independence (CI) test and take into account only the edges above with straight lines:
How many edges do CI tests remove according to a significance level of 0.05? Answer=[Blank 1]
How many edges do CI tests remove according to a significance level of 0.01? Answer=[Blank 2]

Assuming the use of CI tests and considering all edges above with straight and doted lines:
How many edges do CI tests remove according to a significance level of 0.05? Answer=[Blank 3]
How many edges do CI tests remove according to a significance level of 0.01? Answer=[Blank 4]

What are the p-values for the doted lines using the CI test 'cressie_read':
'Smoking' -> 'Coughing': Answer=[Blank 5]
'Smoking' -> 'Car_Accident': Answer=[Blank 6]
'Anxiety' -> 'Fatigue': Answer=[Blank 7]
'Anxiety' -> 'Attention_Disorder': Answer=[Blank 8]
'Allergy' -> 'Fatigue': Answer=[Blank 9]

Use program `bnlearn_ConditionalIndependenceTests.py` to answer the questions above, and provide your answers using the first 4 decimals without rounding.

Blank 1

Blank 2

Blank 3

Blank 4

Blank 5

Blank 6

Blank 7

Blank 8

Blank 9

Question 2

10 points

Using the Lung Cancer dataset, learn the structure using program `bnlearn_StructureLearning.py` by making use of a scoring function.

What is the number of edges induced using the Bayesian Information Criterion (BIC)? Answer=[Blank 1]
What is the number of edges induced using the Akaike Information Criterion (AIC)? Answer=[Blank 2]

Run the program above a few times until you get consistent results (in terms of number of edges) three times in a row. Once that happens, record the information of edges from each of the questions above because those structures will be used in the next question.

Blank 1

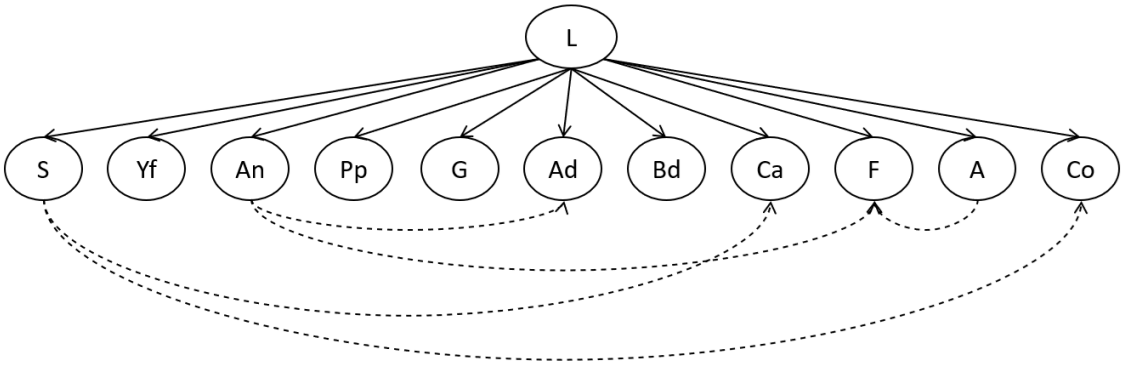
Blank 2

Question 3

50 points

Consider the following Bayes Net structure for the Lung Cancer dataset

S=Smoking
Yf=Yellow_Fingers
An=Anxiety
Pp=Peer_Preassure
G=Genetics
Ad=Attention_Disorder
Bd=Born_an_Even_Day
Ca=Car_Accident
F=Fatigue
A=Allergy
Co=Coughing
L=Lung_cancer



and the following four different variants:

1. Naïve Bayes structure (edges with straight lines above)
2. Extended predefined structure (all edges above with straight and dotted lines)
3. Bayes net structure learnt using the BIC scoring function (output from Question 2)
4. Bayes net structure learnt using the AIC scoring function (output from Question 2)

What is the performance of structure 1? Balanced Accuracy=[Blank 1], KL-Divergence=[Blank 2]
What is the performance of structure 2? Balanced Accuracy=[Blank 3], KL-Divergence=[Blank 4]
What is the performance of structure 3? Balanced Accuracy=[Blank 5], KL-Divergence=[Blank 6]
What is the performance of structure 4? Balanced Accuracy=[Blank 7], KL-Divergence=[Blank 8]

Answering these questions requires two steps:

- (a) train the parameters of your Bayes net as done last week (specifying yourself each config file without CPTs and using `CPT_Generator.py` to generate CPTs automatically), and
- (b) evaluate the learnt structure and parameters using program python `ModelEvaluator.py`.

These steps are important and beneficial for your assignment. So, make sure you understand what is being done here and how. Get in touch with your lecturer or demonstrator if you have any question.

Blank 1

Blank 2

Blank 3

Blank 4

Blank 5

Blank 6

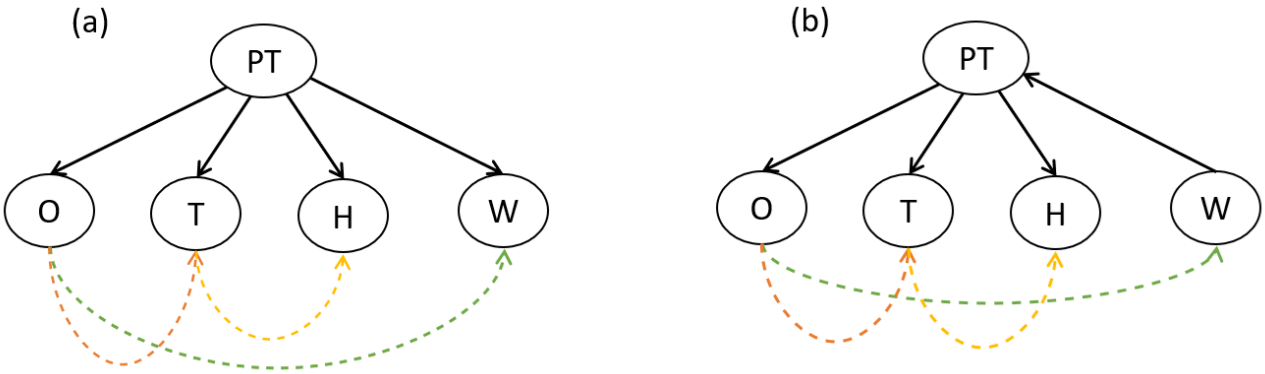
Blank 7

Blank 8

Question 4

5 points

Since Bayesian networks require directed acyclic graphs, it is important to detect whether a graph is indeed acyclic or not. Use the following example graphs to detect the presence of a loop.



Which graph(s) exhibit a loop according to the following code (using the example graph of lecture slide with title "*Depth First Search (DFS) for Detecting Cyclic Networks*"):

```
import BayesNetUtil as bnu
G_lecture_slide= [('A', 'B'), ('A', 'D'), ('B', 'C'), ('C', 'D'), ('D', 'E'),
('E', 'B')]
bnu.has_cycles(G_lecture_slide)
```

A Neither of them

B (a)

C (b)

D Both of them

Question 5

15 points

Which of the following choices is used in the code of this week's workshop?

A Bayesian network training using MLE with Laplace smoothing and inference via Variable Elimination

B Bayesian network training using MLE with Adaptive smoothing and inference via Variable Elimination

Ⓒ Bayesian network training using MLE with Dirichlet priors and inference via Variable Elimination

Ⓓ Bayesian network training using MLE with Laplace smoothing and reasoning via inference by enumeration

Ⓔ Bayesian network training using MLE with Adaptive smoothing and reasoning via inference by enumeration

Ⓕ Bayesian network training using MLE with Dirichlet priors and reasoning via inference by enumeration

Ⓖ Bayesian network training using MLE with Laplace smoothing and inference via other algorithms (not the above)

Ⓗ Bayesian network training using MLE with Adaptive smoothing and reasoning via other algorithms (not the above)

Ⓘ Bayesian network training using MLE with Dirichlet priors and reasoning via other algorithms (not the above)