CMP9794M Advanced Artificial Intelligence – Workshop Week 3

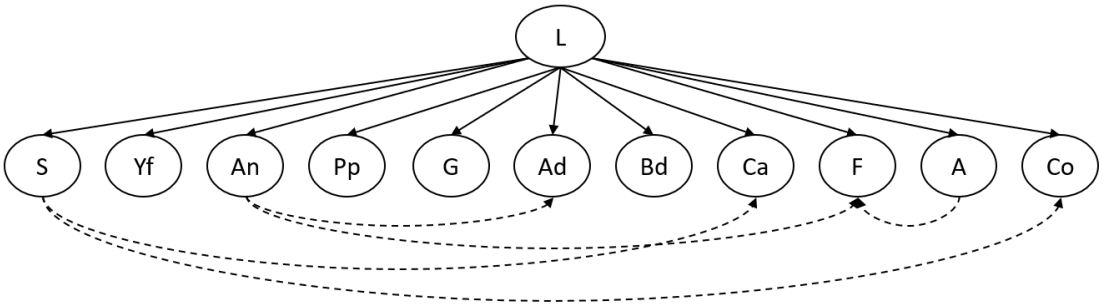
UNIVERSITY OF LINCOLN

Summary: In this workshop you will carry out calculations related to structure learning in Bayes nets. In particular, those related to conditional independent tests, log-likelihood and Bayesian information criterion. In addition, this workshop addresses the evaluation of probabilistic models. Look at the update class diagram in file `workshop-w3/doc/CMP9794_BayesNets_ClassDiagram.pdf`. From Blackboard, download the Zip file from the workshop materials of this week and install the following dependencies: `pip install causal-learn; networkx`

Task 1: **Conditional Independence Tests**

Consider the following Bayes Net structure for the Lung Cancer dataset used in the workshop of week 1 (note that the long names of random variables are the 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



a. Carry out conditional independence tests for the dashed arrows above as follows:

- `python ConditionalIndependence.py ..\data\lung_cancer-train.csv "I(Smoking,Coughing|Lung_cancer)"`
- `python ConditionalIndependence.py ..\data\lung_cancer-train.csv "I(Smoking,Car_Accident|Lung_cancer)"`
- `python ConditionalIndependence.py ..\data\lung_cancer-train.csv "I(Anxiety,Fatigue|Lung_cancer)"`
- `python ConditionalIndependence.py ..\data\lung_cancer-train.csv "I(Anxiety,Attention_Disorder|Lung_cancer)"`
- `python ConditionalIndependence.py ..\data\lung_cancer-train.csv "I(Allergy,Fatigue|Lung_cancer)"`

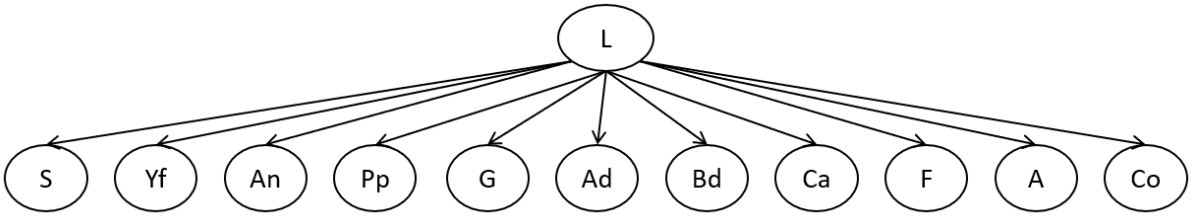
a. What edges above would you remove according to a significance level of 0.05?
b. What edges above would you remove according to a significance level of 0.01?

The program above uses Chi-Square tests by default. In your own time, you can verify your results with the G^2 test by setting the value of binary variable `chi_square_test` to False.

Task 2: **Scoring functions and model performance**

Using the following structures together with the Lung Cancer data,

a. What is the Log-Likelihood (LL) of a Naïve Bayes model (diagram below)?
b. What is the Bayesian Information Criterion (BIC) of the same data and model?
c. What is the performance of the same model but on test data?



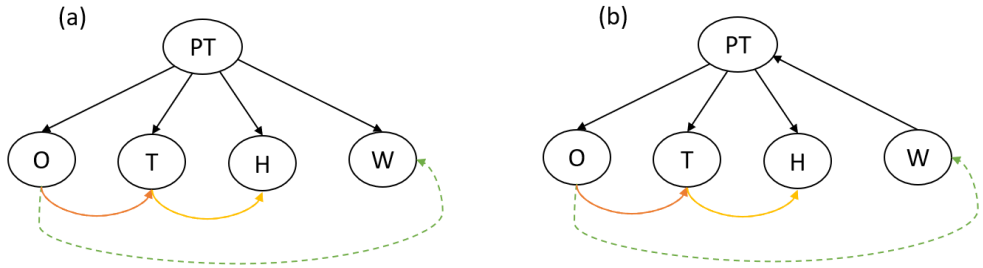
To answer these questions, run the following program:

- `python ModelEvaluator.py ..\config\config-lungcancer.txt ..\data\lung_cancer-train.csv ..\data\lung_cancer-test.csv`

Open that program and try to understand its implementation and ask questions to your lecturer.

Task 3: **Detecting loops in graphs/networks**

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



Run the following code as an example using the graph shown in slide 25 of this week's lecture:

```
import BayesNetUtil as bnu

G_slide25 = [('A', 'B'), ('A', 'D'), ('B', 'C'), ('C', 'D'), ('D', 'E'),
('E', 'B')]

bnu.has_cycles(G_slide25)
```

What results do you get using diagrams (a) and (b)?

Task 4: **Homework**

4.1 What is the Log-Likelihood (LL), Bayesian Information Criterion (BIC) and performance of a Bayesian Network using the structure shown in task 1?

4.2 Use the programs above to implement the Min-Max-Hill Climbing algorithm discussed in the previous lecture.