27 March 2023

UNIVERSITY OF GLOUCESTERSHIRE

CT4031
Maths for Data Science

Week7 - Probability Part 3

How to estimate probabilities from data?

For categorical data is simple:

Class: $P(C) = N_c/N$

Single:

- P(Evade=No) = 7/10
- P(Evade= Yes) = 3/10

"Given"

- P(Status=Married|Evade= No) = 4/7
- P(Refund=Yes|Evade= Yes)=0

Tid	**Refund**	**Marital Status**	**Taxable Income**	**Evade**
1	Yes	Single	125K	**No**
2	No	Married	100K	**No**
3	No	Single	70K	**No**
4	Yes	Married	120K	**No**
5	No	Divorced	95K	**Yes**
6	No	Married	60K	**No**
7	Yes	Divorced	220K	**No**
8	No	Single	85K	**Yes**
9	No	Married	75K	**No**
10	No	Single	90K	**Yes**

How to estimate probabilities from data?

Tid	**Refund**	**Marital Status**	**Taxable Income**	**Evade**
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Formula:

$$P(A_i \mid c_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(A_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

For (Income, Class=No):

- If Class=No
  - Mean μ_{ij} = 104
  - Variance σ_{ij}^2 = 2975

$$P(Income = 120 \mid No) = \frac{1}{\sqrt{2\pi}(54.54)} e^{-\frac{(120 - 104)}{2(2975)}} = 0.007$$ =>7%

Bayesian Classifiers

- Naïve Bayes Classifier

Bayesian Classifiers - Example

$$X = (\text{Refund} = \text{No}, \text{Married}, \text{Income} = 120\text{K})$$

naive Bayes Classifier:

P(Refund=Yes|No) = 3/7
P(Refund=No|No) = 4/7
P(Refund=Yes|Yes) = 0
P(Refund=No|Yes) = 1
P(Marital Status=Single|No) = 2/7
P(Marital Status=Divorced|No)=1/7
P(Marital Status=Married|No) = 4/7
P(Marital Status=Single|Yes) = 2/7
P(Marital Status=Divorced|Yes)=1/7
P(Marital Status=Married|Yes) = 0

For taxable income:
If class=No: sample mean=110
sample variance=2975
If class=Yes: sample mean=90
sample variance=25

P(X|Class=No) = P(Refund=No|Class=No)
$\times$ P(Married| Class=No)
$\times$ P(Income=120K| Class=No)
= 4/7 $\times$ 4/7 $\times$ 0.0072 = 0.0024

P(X|Class=Yes) = P(Refund=No| Class=Yes)
$\times$ P(Married| Class=Yes)
$\times$ P(Income=120K| Class=Yes)
= 1 $\times$ 0 $\times$ 1.2 $\times$ 10^{-9} = 0

What is Model

- A representation of another object
- make predictions, test assumptions, and solve problems
- perform useful functions.
  - a map is a model of a location, which can help us get from place to place.
  - a bus timetable is a model of where buses should appear
  - a company’s financial reports are a model of one aspect of the company

Different models

- Models can be:
  - Deterministic
  - Probabilistic

Different models

- Models are used to make predictions, test assumptions, and solve problems.
- Models can be:
  - Deterministic
    - Does not include elements of randomness.
    - Every time, the model with the same conditions produces the same results.
  - Probabilistic
    - Includes elements of randomness.
    - Every time, the model is likely to get different results, even with the same conditions.

Different models

- Deterministic models
  - Basic programing
  - Basic functions
  - Code that sums two numbers
  - The equation to calculate the average speed of a car
- Probabilistic models
  - Prediction
  - Image recognition
  - Code that diagnosis a patient with cancer

Different models

- Customers arrive to use a cash machine every two minutes on average.
- Customers take 2 minutes to use the machine on average.
- What is the probability that a customer has to wait 3 minutes or more?

Different models

- A deterministic model:
  - Uses the average gap between customers and the average time of usage
  - The model assumes that someone arrives exactly every two minutes and uses the machine for exactly two minutes, so there is never any waiting time.
- What do you think of this model?

Probabilistic models

- Includes elements of randomness.
- Every time, the model is likely to get different results, even with the same conditions.
- Prediction
- Example
  - Code that diagnosis a patient with cancer
  - Queueing models
  - Markov chains
  - Baysian Belief Network

Probabilistic models

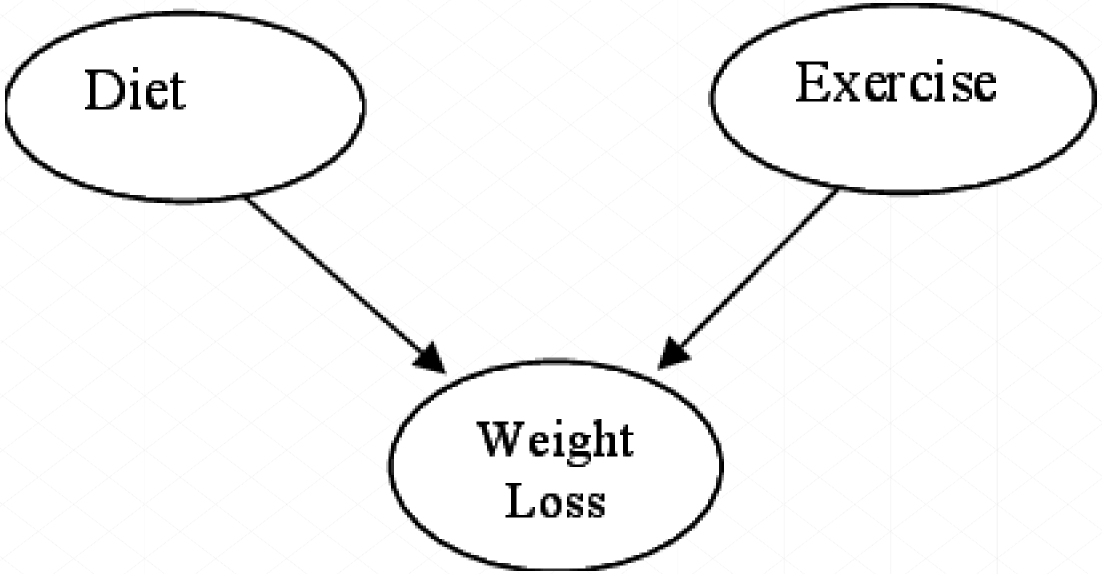
- A probabilistic model might keep the time of use at the machine as 2 minutes for each person, but include random arrival times.
- We can think of 15 different arrival times in a 30 minute period, for example, 2 4 5 5 10 11 12 15 16 19 20 24 29 29 29.
- What is the difference?

Probabilistic models

- A probabilistic model might keep the time of use at the machine as 2 minutes for each person, but include random arrival times.
- We can think of 15 different arrival times in a 30 minute period, for example, 2 4 5 5 10 11 12 15 16 19 20 24 29 29 29.
- The average waiting time is still 2, however, it might have different probabilities depending on the users before me, the time that I am arriving, etc.

Bayesian Network

- What is this?



Introduction to Bayesian network

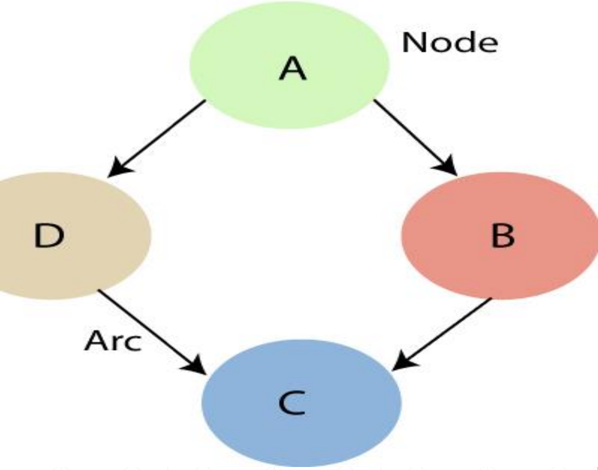
- Bayesian networks are a probabilistic graphical model comprised of nodes and directed edges.
- Enables modelling of beliefs in the states of the variables in terms of probabilities.
- Bayesian networks take events that occurred and are able to predict the likelihood that any one of the known causes was a contributing factor.

Introduction to Bayesian network

- Also known as a Bayes network, Bayes net, belief network, or decision network
- A probabilistic graphical model
- Represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG).
- Ideal for taking an event that occurred and predicting the likelihood that any one of several possible known causes was the contributing factor.
- **Example:**
  - Given symptoms, the network can be used to compute the probabilities of the presence of various diseases.

Introduction to Bayesian network

- Build models from data and/or expert opinion using directed acyclic graph (DAG)
- **Nodes**
  - random variables (continuous or discrete).
- **Directed Links**
  - causal relationship or conditional probabilities between random variables
  - A, B, C, and D are random variables represented by the nodes of the network graph.
  - If we are considering node B, which is connected with node A by a directed arrow, then node A is called the parent of Node B.
  - Node C is independent of node A.
- **Conditional Probability Table (CPT):**
  - The values in CPT are based on the experimental data or observations



Introduction to Bayesian network

- Based on Joint probability distribution and conditional probability.
- Each node in the Bayesian network has condition probability distribution P(Xi |Parent(Xi)), which determines the effect of the parent on that node.
- If we have variables x1, x2, x3,....., xn, then the probabilities of a different combination of x1, x2, x3.. xn, are known as Joint probability distribution.
- P[x1, x2, x3,....., xn], it can be written as the following way in terms of the joint probability distribution.
- = P[x1| x2, x3,....., xn]P[x2, x3,....., xn]
- = P[x1| x2, x3,....., xn]P[x2|x3,....., xn]....P[xn-1|xn]P[xn].
- In general for each variable Xi, we can write the equation as:
- P(Xi|Xi-1,........., X1) = P(Xi |Parents(Xi))

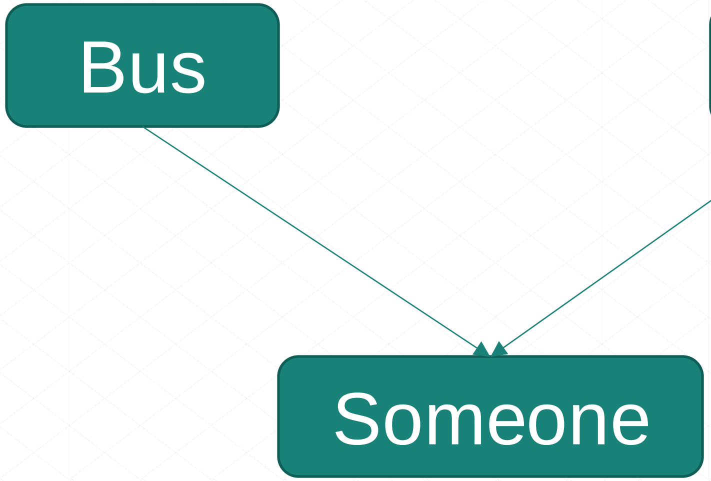
Bayesian Network: Example 1

- Someone catches a bus every morning to the train station
- After that, this someone catches a train to work
- Some mornings the bus is late and sometimes the train is late.
- The bus is late more often than the train
- If either mode of transport should be late it does not necessarily mean that Someone is late for work.

Example (cont.)

P(Bus)	
Late 0.3	On Time 0.7

P(Train)	
Late 0.1	On Time 0.9



Example (cont.)

Someone is late, conditional probability

P(Someone | Bus, Train)			
Bus	**Train**	**Late**	**On Time**
On time	On time	0.01	0.99
On time	Late	0.90	0.10
Late	On time	0.20	0.80
Late	Late	0.90	0.10

Example (cont.)

Bus	Train	Someone	P(Someone|Bus, Train)	P(Bus)	P(Train)	Product
OT	OT	OT	0.99	0.70	0.90	0.6237
OT	L	OT	0.10	0.70	0.10	0.0007
L	OT	OT	0.80	0.30	0.90	0.216
L	L	OT	0.10	0.30	0.10	0.0003
OT	OT	L	0.01	0.70	0.90	0.0063
OT	L	L	0.90	0.70	0.10	0.063
L	OT	L	0.20	0.30	0.90	0.054
L	L	L	0.90	0.30	0.10	0.027

Example (cont.)

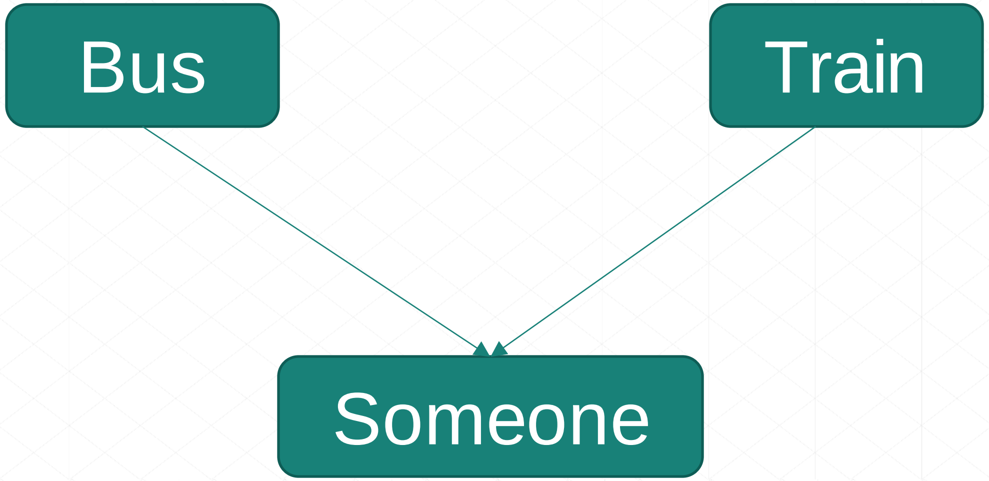
Bus	Train	Someone	P(Someone|Bus, Train)	P(Bus)	P(Train)	Product
OT	OT	OT	0.99	0.70	0.90	0.6237
OT	L	OT	0.10	0.70	0.10	0.0007
L	OT	OT	0.80	0.30	0.90	0.216
L	L	OT	0.10	0.30	0.10	0.0003
OT	OT	L	0.01	0.70	0.90	0.0063
OT	L	L	0.90	0.70	0.10	0.063
L	OT	L	0.20	0.30	0.90	0.054
L	L	L	0.90	0.30	0.10	0.027

Given that Someone is late, what is the probability of this being due to a bus?

Example (cont.)

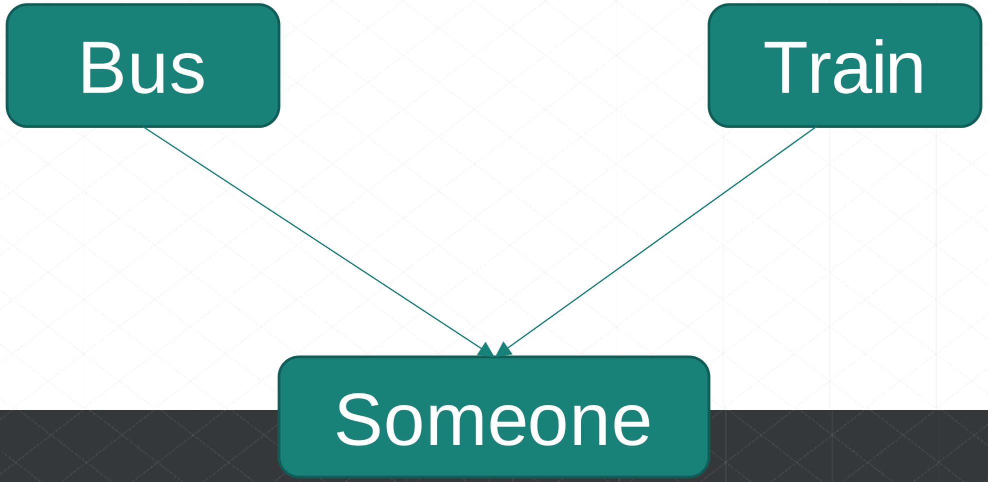
P(Bus)	
Late 0.3	On Time 0.7

P(Train)	
Late 0.1	On Time 0.9



Example (cont.)

- Given that Someone is late, what is the probability of this being due to a bus?
- P(Bus=Late | Someone=Late)
- Using the Bayesian network, the relationship is calculated based on
  - P(Bus=Late & Someone=Late) / P(Someone=Late)



Example (cont.)

P(Bus=Late & Someone=Late) **/** P(Someone=Late) =>

P(Bus=Late & Someone=Late) = 0.054+0.027=0.081

Bus	Train	Someone	P(Someone|Bus, Train)	P(Bus)	P(Train)	Product
L	OT	L	0.20	0.30	0.90	0.054
L	L	L	0.90	0.30	0.10	0.027

Example (cont.)

P(Bus=Late & Someone=Late) **/** P(Someone=Late) =>

P(Bus=Late & Someone=Late) = 0.054+0.027=0.081

P(Someone=Late) = 0.0063+0.063+0.054+0.027 = 0.150

Bus	Train	Someone	P(Someone |Bus, Train)	P(Bus)	P(Train)	Product
OT	OT	**L**	0.01	0.70	0.90	**0.0063**
OT	L	**L**	0.90	0.70	0.10	**0.063**
L	OT	**L**	0.20	0.30	0.90	**0.054**
L	L	**L**	0.90	0.30	0.10	**0.027**

Example (cont.)

P(Bus=Late & Someone=Late) **/** P(Someone=Late) =>

P(Bus=Late & Someone=Late) = 0.054+0.027=0.081

P(Someone=Late) = 0.0063+0.063+0.054+0.027 = 0.150

P(Bus=Late & Someone=Late) **/** P(Someone=Late) = 0.081/0.150 = 0.54

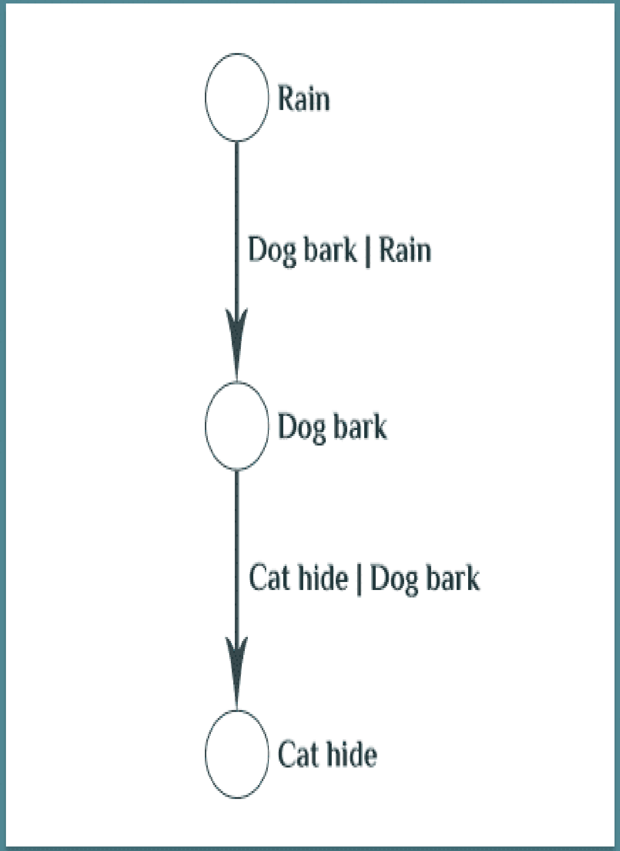
Bayesian Network: Example 2

- Imagine you have a dog that really enjoys barking at the window whenever it's raining outside. Not necessarily every time, but still quite frequently. You also own a sensitive cat that hides under the couch whenever the dog starts barking. Again, not always, but she tends to do it often.

- To continue the example above, if you're outside your house and it starts raining, there will be a high probability that the dog will start barking. This, in turn, will increase the probability that the cat will hide under the couch. You see how information about one event (rain) allows you to make inferences about a seemingly unrelated event (the cat hiding under the couch).

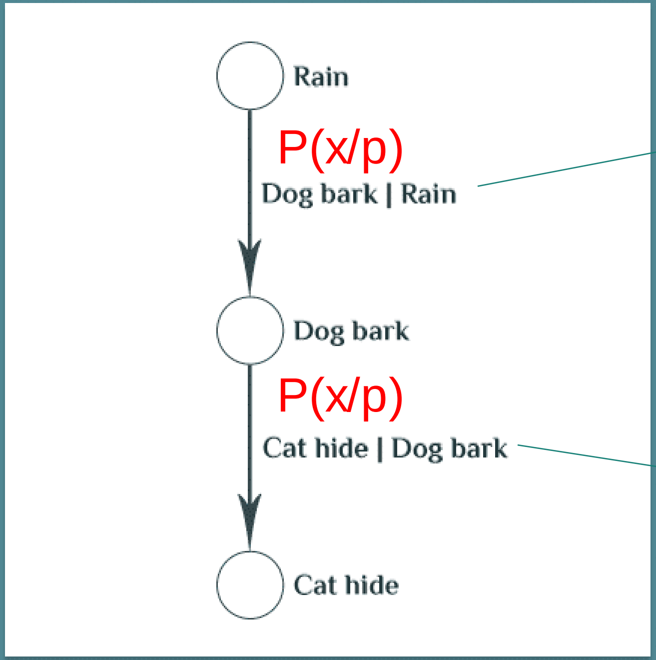
- You can also make the inverse inference. If you see the cat hiding under the couch, this will increase the probability that the dog is currently barking. And that, in turn, will increase the probability that it's currently raining.

(BAYES NETWORK EXAMPLE)



Example (cont.)

(BAYES NETWORK EXAMPLE)



(RAIN/BARK DISTRIBUTION)

	Rains	Doesn't rain	
Dog barks	9/48	18/48	27/48
Dog doesn't bark	3/48	18/48	21/48
	12/48	36/48	48/48

(BARK/HIDE DISTRIBUTION)

	Cat hides	Cat doesn't hide	
Dog barks	18/48	9/48	27/48
Dog doesn't bark	7/48	14/48	21/48
	25/48	23/48	48/48

Bayesian Network: Example 3

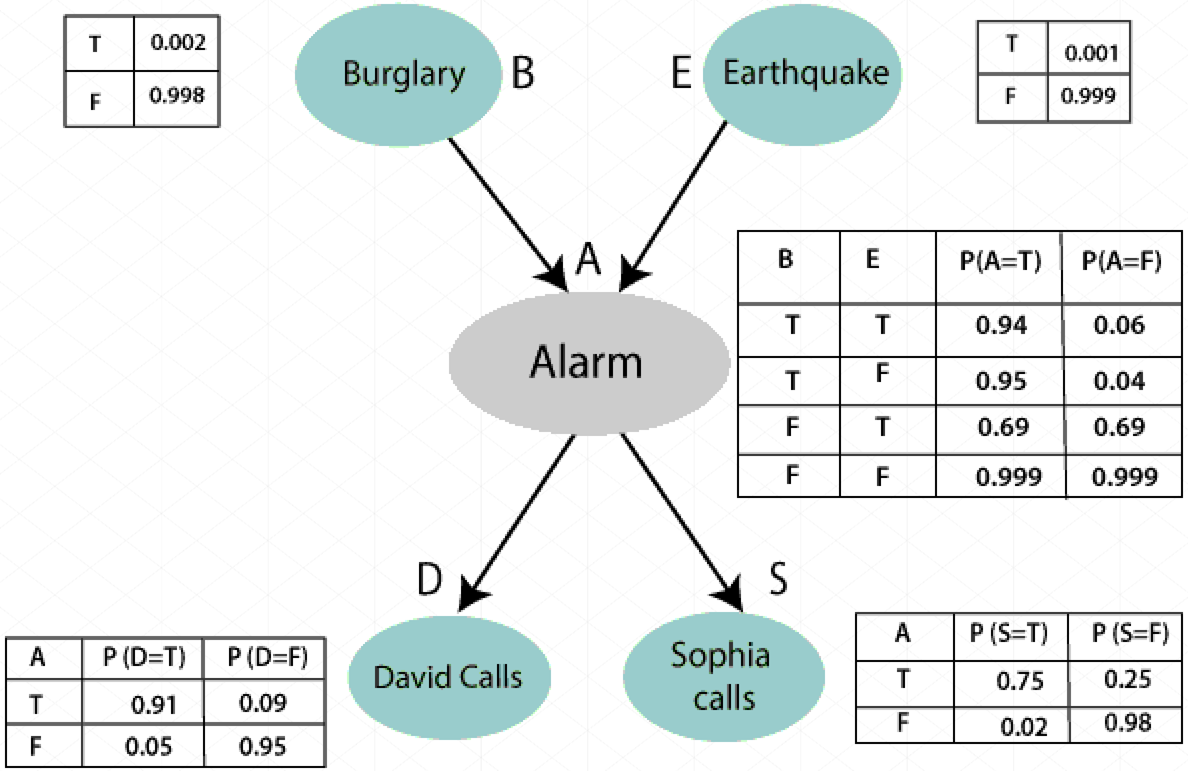
- Example: Harry installed a new burglar alarm at his home to detect burglary. The alarm reliably responds at detecting a burglary but also responds for minor earthquakes. Harry has two neighbors David and Sophia, who have taken a responsibility to inform Harry at work when they hear the alarm. David always calls Harry when he hears the alarm, but sometimes he got confused with the phone ringing and calls at that time too. On the other hand, Sophia likes to listen to high music, so sometimes she misses to hear the alarm. Here we would like to compute the probability of Burglary Alarm.

- **Problem:**

  - **Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and David and Sophia both called the Harry.**

Example (cont.)

- List of all events occurring in this network:
  - Burglary (B)
  - Earthquake(E)
  - Alarm(A)
  - David Calls(D)
  - Sophia calls(S)



T	0.002
F	0.998

T	0.001
F	0.999

B	E	P(A=T)	P(A=F)
T	T	0.94	0.06
T	F	0.95	0.04
F	T	0.69	0.69
F	F	0.999	0.999

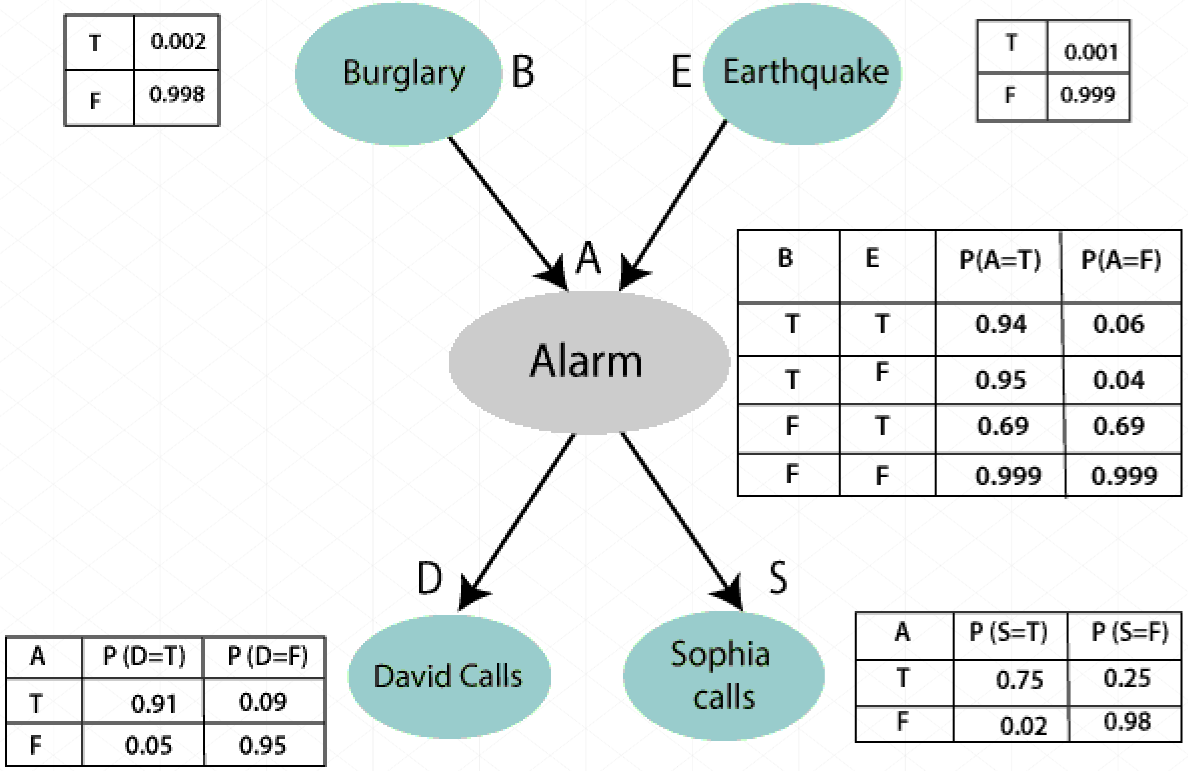
A	P (D=T)	P (D=F)
T	0.91	0.09
F	0.05	0.95

A	P (S=T)	P (S=F)
T	0.75	0.25
F	0.02	0.98

Example (cont.)

- Conditional probability table for Alarm A:
- The Conditional probability of Alarm A depends on Burglar and earthquake:

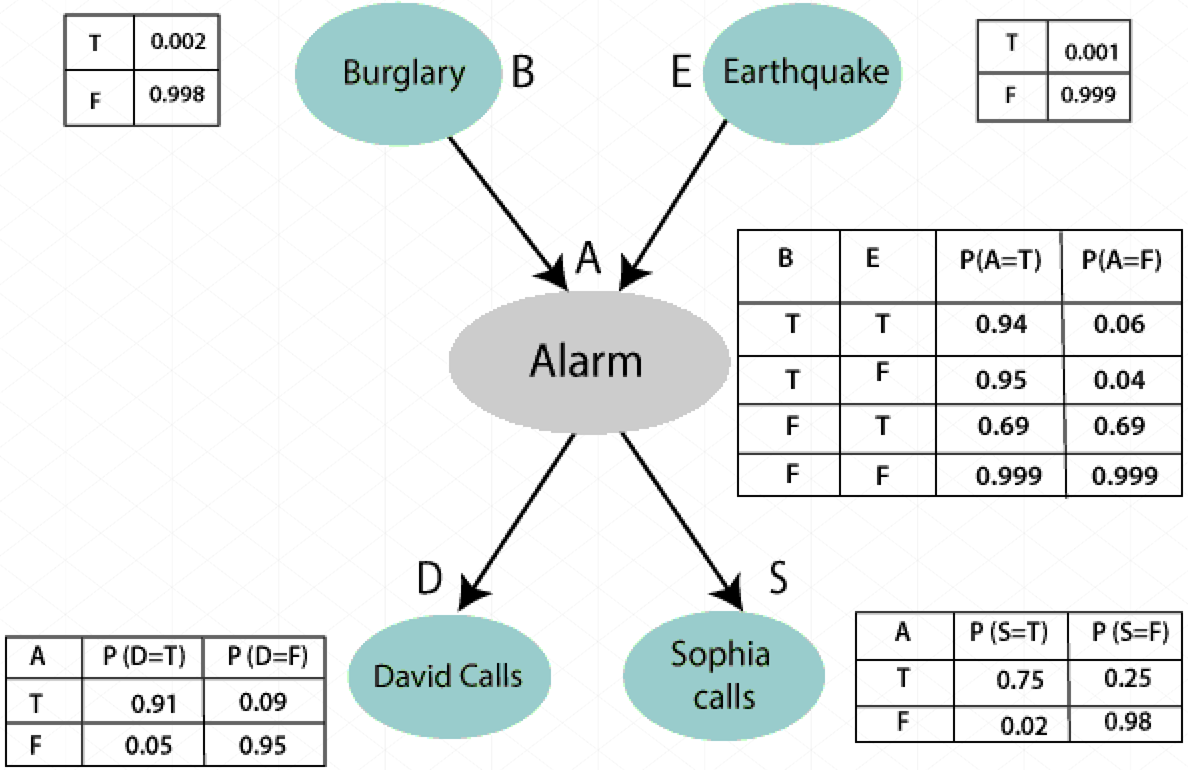
B	E	P(A= True)	P(A= False)
True	True	0.94	0.06
True	False	0.95	0.04
False	True	0.31	0.69
False	False	0.001	0.999



Example (cont.)

- Conditional probability table for David Calls:
- The Conditional probability of David that he will call depends on the probability of Alarm.

A	P(D= True)	P(D= False)
True	0.91	0.09
False	0.05	0.95

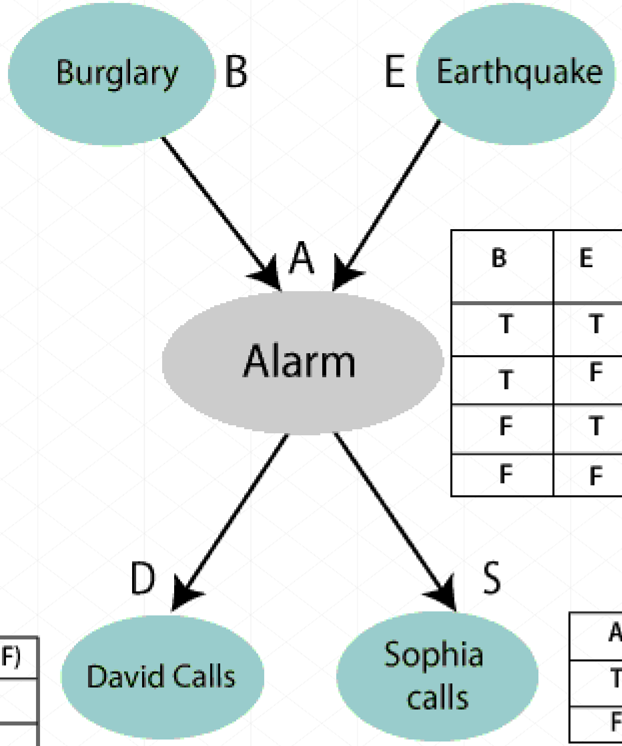


❑ Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and David and Sophia both called the Harry.

- ❑ Conditional probability table for Sophia Calls:
- ❑ The Conditional probability of Sophia that she calls is depending on its Parent Node "Alarm.“

A	P(S= True)	P(S= False)
True	0.75	0.25
False	0.02	0.98

- ❑ From the formula of joint distribution, we can write the problem statement in the form of probability distribution:
- ❑ P(S, D, A, ¬B, ¬E) = P (S|A) *P (D|A)*P (A|¬B ^ ¬E) *P (¬B) *P (¬E).
- ❑ = 0.75* 0.91* 0.001* 0.998*0.999
- ❑ = 0.00068045.
- ❑ Hence, a Bayesian network can answer any query about the domain by using Joint distribution.



T	0.002
F	0.998

T	0.001
F	0.999

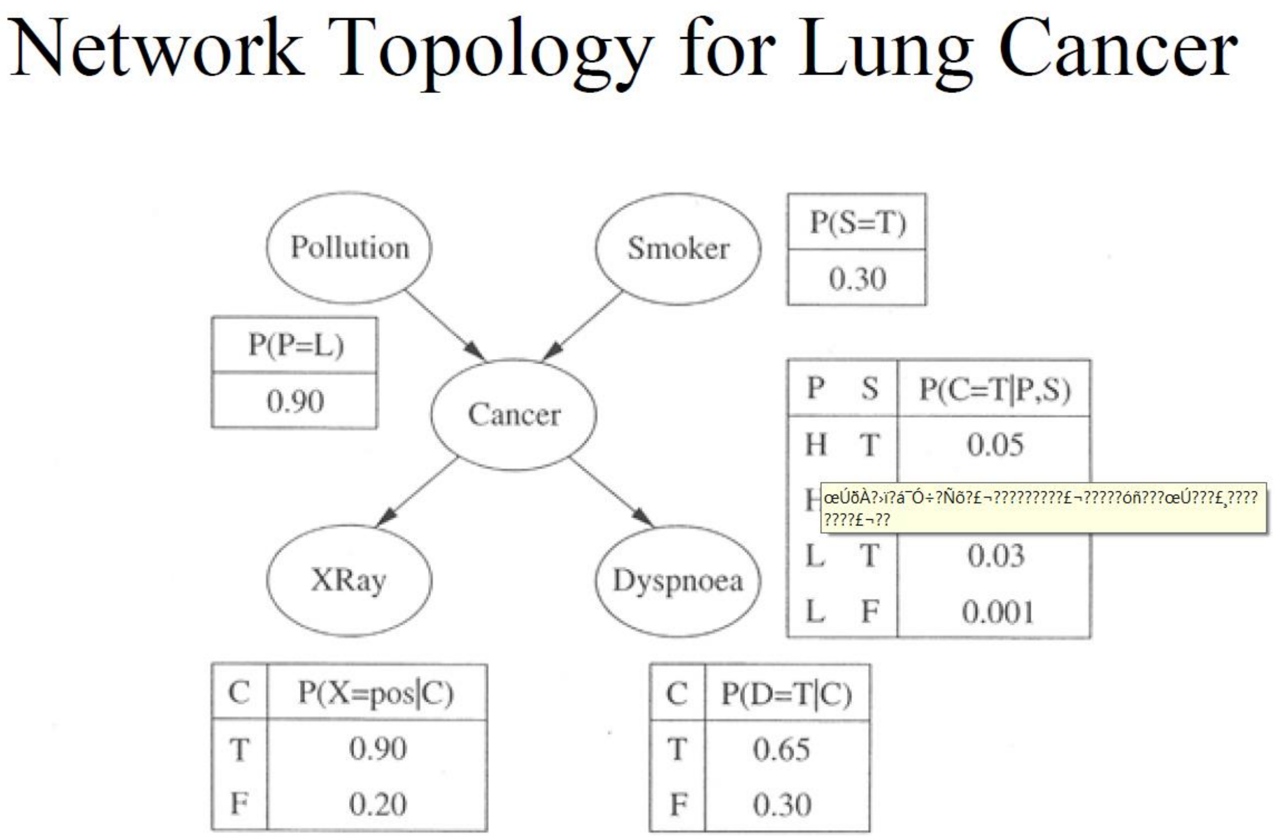
B	E	P(A=T)	P(A=F)
T	T	0.94	0.06
T	F	0.95	0.04
F	T	0.69	0.69
F	F	0.999	0.999

A	P (D=T)	P (D=F)
T	0.91	0.09
F	0.05	0.95

A	P (S=T)	P (S=F)
T	0.75	0.25
F	0.02	0.98

Bayesian Network: Example 4

- Can you please what this baysian model is trying to predict?
- What data has been provided?
- Who is the parent node of Cancer?
- Find the probability of cancer when the patient is smoker , living in clean environment, and x ray shows the sign of cancer, and no Dyspnoea.



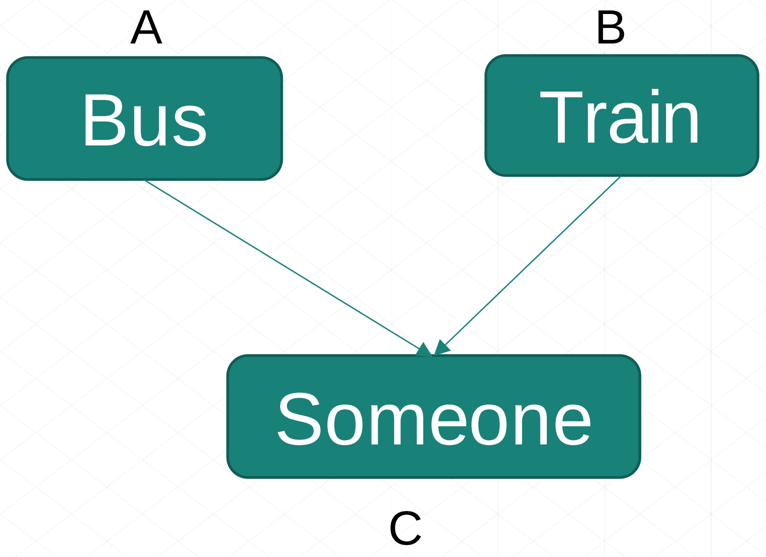
Bayesian Network

- Different Connections
- Converging
- Diverging
- Serial
    - Forward
    - Backward

Bayesian Network

Convergence

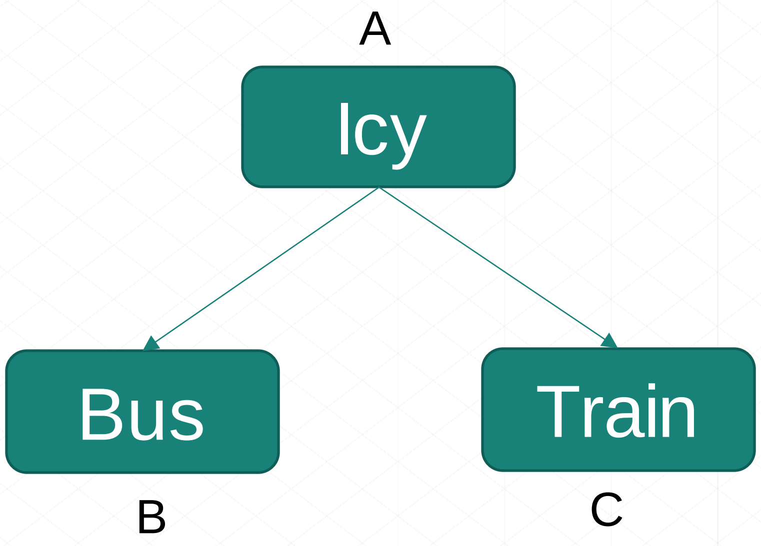
- A and B are independent unless information about C or a descendent of C has been received.



Bayesian Network

Divergence

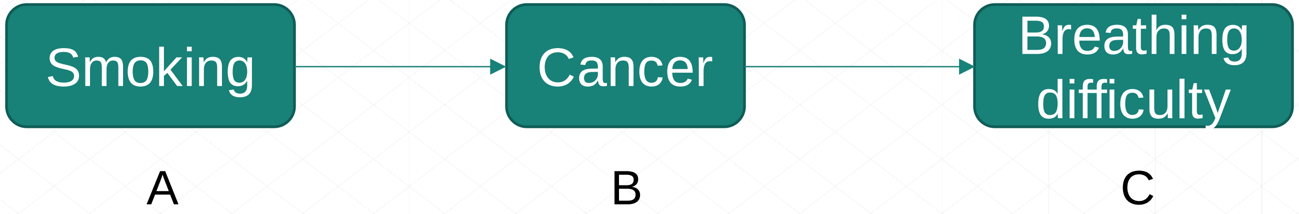
- B and C are dependent unless evidence about A is received.



Bayesian Network

Serial (Forward or Backward)

- A and C are dependent and become independent upon receiving evidence about B.



Bayesian Network

Task – Provide real-world examples of:

- Converging
- Diverging
- Serial

Practice Activity 1

Calculate the probability that alarm has sounded due to burglary but there is no an earthquake occurred, and neither David and nor Sophia called the Harry.

Practice Activity 2

Calculate the probability that alarm has sounded, but there is an earthquake occurred but not the burglary, and only David called the Harry.

In summary

- Bayesian network is a DAG (directed acyclic graph) in which each node represents a variable.
- Each variable has an associated conditional probability table.
- The network and table provide a joint probability distribution for the variables.

Post-sessional work

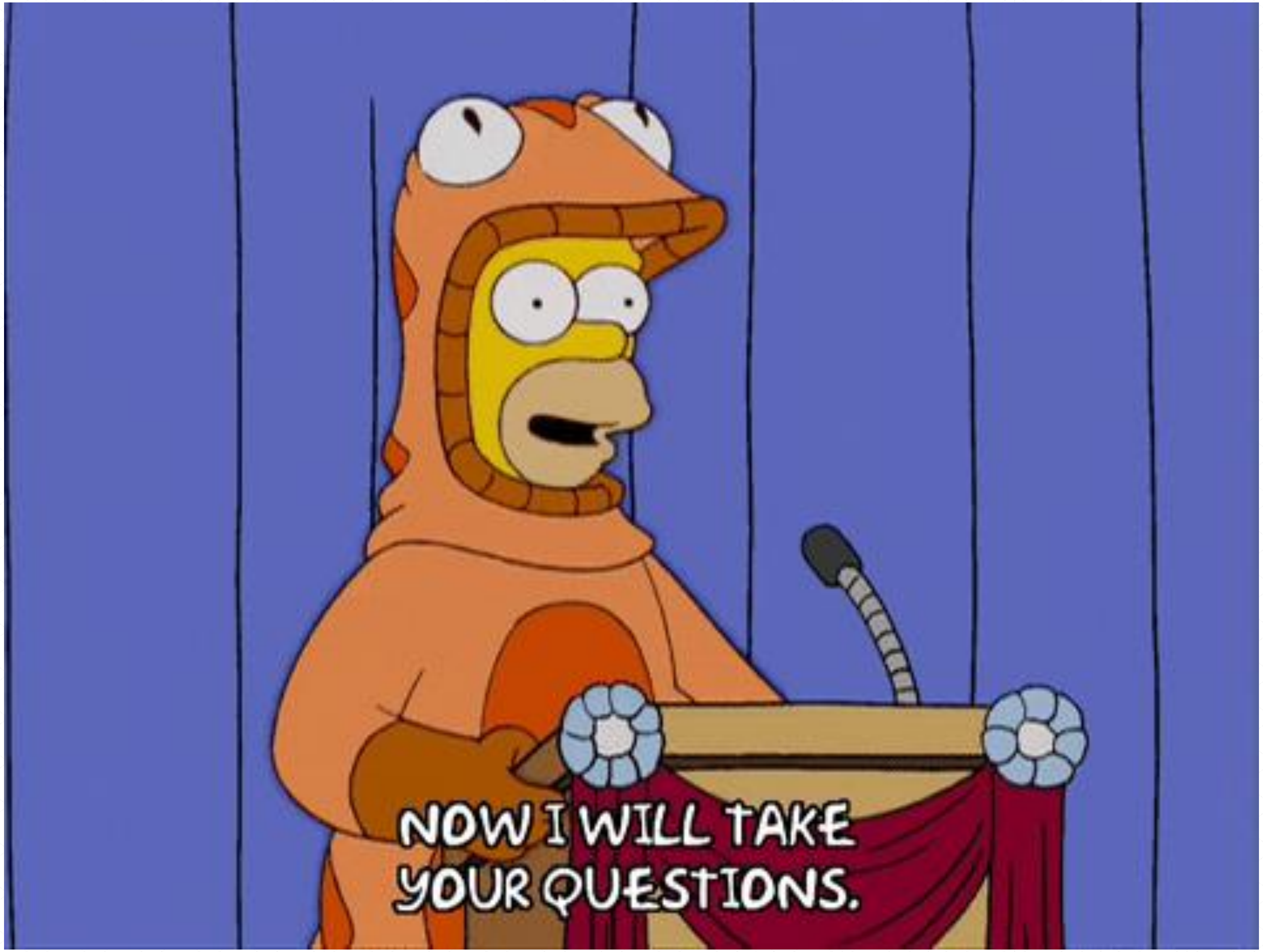
- Work on the assignment.
- We will be taking assignment-related questions next week in the first 15 minutes of the week 8 session.

References

- Callan, R. (2003). *Artificial Intelligence*. Part-3. Palgrave Macmillan
- B. Illowsky and S. Dean, 'The Exponential Distribution', Jul. 2013, Accessed: Jul. 15, 2021. [Online]. Available: https://opentextbc.ca/introstatopenstax/chapter/the-exponential-distribution/

Next Session !

- Probability Part 4


NOW I WILL TAKE
YOUR QUESTIONS.