27 March 2023

UNIVERSITY OF GLOUCESTERSHIRE

CT4031
Maths for Data Science

Week 5 – Probability Part 2

Uncertainty: Probability

Conditional Independence

- Two variables are independent if the state of one does not influence the state of the other.

- In this case: $P(A|B, C) = P(A|B)$ or $P(A|C)$

Conditional Independence

- Two variables are independent if the state of one does not influence the state of the other.

- For example:

	Literacy	
T-shirt	Yes	No
Small	0.32	0.68
Large	0.35	0.65

		Literacy	
Age	T-shirt	Yes	No
<5	Small	0.3	0.7
<5	Large	0.3	0.7
>5	Small	0.4	0.6
>5	Large	0.4	0.6

Mutually exclusive vs Independent

- A **mutually exclusive** event can simply be defined as a situation when two events cannot occur at same time whereas **independent** event occurs when one event remains

- **Mutually exclusive** events are those that cannot happen simultaneously, whereas **independent** events are those whose probabilities do not affect one another (unaffected by the occurrence of the other event).

Mutually exclusive probability

- AKA – Disjoint event
  - Probability of raining and not raining
  - Probability of having COVID and not having COVID
  - Probability of passing the module and not passing the module

Conditional Dependency

- Conditional probability is the probability of an event happening, given that another event has also happened.
- Examples:
  - Probability of raining tomorrow given that it is sunny today
  - Probability of having COVID given that you have 2 out of 3 symptoms.
  - Probability of passing the module given that you have been doing all post-sessional work.

Conditional Dependency

- For that, we use Bayes Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Conditional Dependency

- Yet another example
- Suppose a new Ransomware has been developed.
- The Ransomware is 99% identifiable, that is the proportion of devices who are correctly identified by an antivirus as having the malware.
- Therefore, there is a 1% false positive rate.
- Suppose that 0.5% of devices at UoG have the Ransomware.
- What is the probability that a randomly selected device at UoG being flagged as having the Malware, actually having the Malware?

Conditional Dependency

- What is the probability that a randomly selected device at UoG being flagged as having the Malware, actually having the Malware?

$$P(Malware \mid AV+) = \frac{P(AV+ \mid Malware)P(Malware)}{P(AV+)}$$

Conditional Dependency

- The probability of being positive is equals to the probability of a positive flagged by the Antivirus being a Malware PLUS the probability of a FALSE POSITIVE, hence:

$$P(Malware \mid AV+) = \frac{P(AV + \mid Malware)P(Malware)}{P(AV+)}$$

$$= \frac{P(AV + \mid Malware)P(Malware)}{P(AV + \mid Malware)P(Malware) + P(AV + \mid NotMalware)P(NotMalware)}$$

Conditional Dependency

- The Ransomware is 99% identifiable, that is the proportion of devices who are correctly identified by an antivirus as having the malware.
  - Therefore, there is a 1% false positive rate.
- Suppose that 0.5% of devices at UoG have the Ransomware.
  - Therefore, there are 95% of devices not having the Ransomware

$$= \frac{P(AV+|\,Malware)P(Malware)}{P(AV+|\,Malware)P(Malware) + P(AV+|\,NotMalware)P(NotMalware)}$$

$$= \frac{0.99\ x\ 0.005}{0.99\ x\ 0.005\ + 0.01\ x\ 0.995}$$

= 33%

In summary

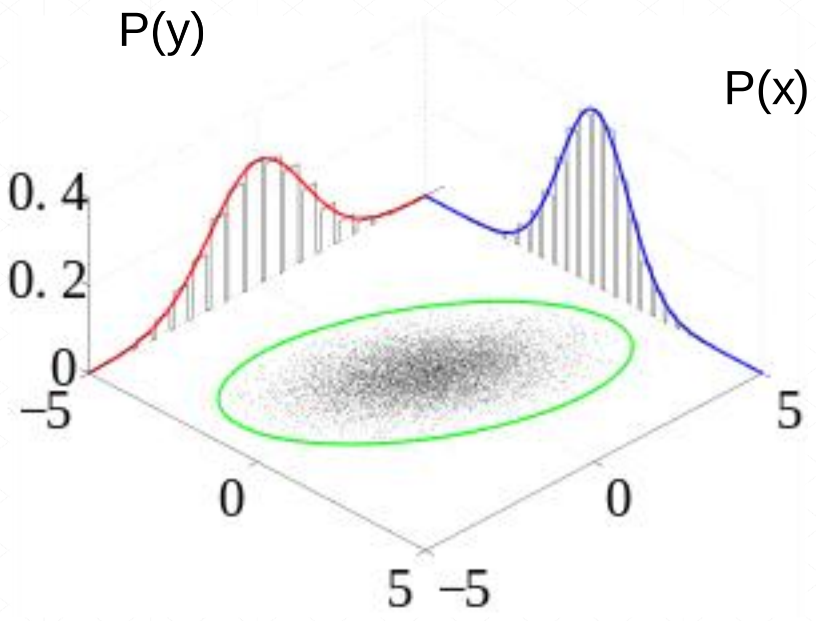
- Events can be:
  - Independent — Each event is not effected by previous or future events.
  - Dependent — An event is affected by other events
  - Mutually Exclusive — Events can't happen at the same time

Joint probability distribution

- Probability quantifies the uncertainty of the outcomes of a random variable.
- It is easy to understand the probability for a single variable.
- HOWEVER, in Data Science, Machine Learning and AI, we often have many variables that interact in complex ways.
- HENCE, Joint probability distribution!
  - Joint probability is the probability of two events occurring simultaneously.
  - Example: the probability of a X=1 and Y=2,

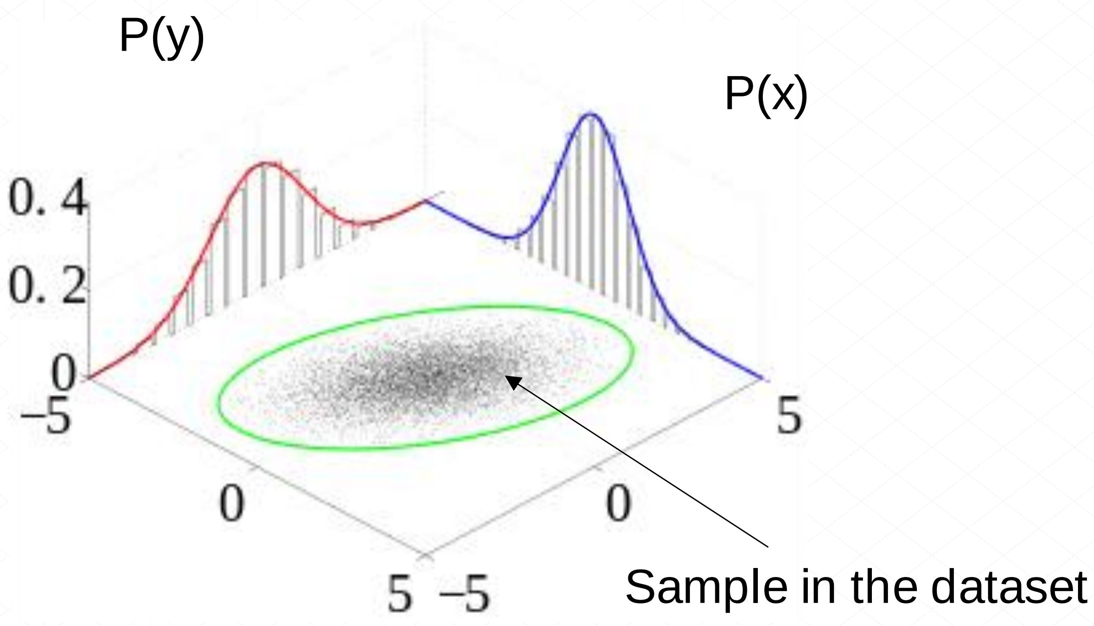
Joint probability distribution

- That works very well for functions and relationship between different features/attributes in a computational problem:



Joint probability distribution

- That works very well for functions and relationship between different features/attributes in a computational problem:



Chain Rule

- A way of factoring a joint distribution into a product of conditional probabilities.

$$P(A, B, C) = P(A|B, C) \times P(B|C) \times P(C)$$

Chain Rule

- A way of factoring a joint distribution into a product of conditional probabilities.

$$P(A, B, C) = P(A|B, C) \times P(B|C) \times P(C)$$

- And that can be generalised for N elements

$$P(A1, A2, ..., An) = P(A1| A2, ..., An)\ P(A2| A3, ..., An)\ P(An\text{-}1|An)\ P(An)$$

Chain Rule

- How can this be used in data science?

Chain Rule

- How can this be used in data science?
- Classification!

Bayesian Classifiers

- Approach: compute the posterior probability $P(C \mid A_1, A_2, \ldots, A_n)$ for all values of C using the Bayes theorem

$$\boxed{P(C \mid A_1 A_2 \ldots A_n) = \frac{P(A_1 A_2 \ldots A_n \mid C) P(C)}{P(A_1 A_2 \ldots A_n)}}$$

- Optimization: how to optimize the results? And why?

Bayesian Classifiers

- Approach: compute the posterior probability $P(C \mid A_1, A_2, \ldots, A_n)$ for all values of C using the Bayes theorem

$$\boxed{P(C \mid A_1 A_2 \ldots A_n) = \frac{P(A_1 A_2 \ldots A_n \mid C) P(C)}{P(A_1 A_2 \ldots A_n)}}$$

- Optimization: choose value that maximizes the chances for C, to improve the accuracy of the classifier.

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

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?

What about for continuous numerical data?

What is the probability of:

- P(Income=120|Evade=No)?
  - Is 1/10 correct?

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?

What about for continuous numerical data?

What is the probability of:

- P(Income=120|Evade=No)?
  - Is 1/10 correct?
  - Are you sure?

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?

What about for continuous numerical data?

What is the probability of:

- P(Income=120|Evade=No)?
  - Is 1/10 correct?
  - Are you sure?
  - Think again!

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?

To calculate that, we need to assume a Normal distribution.

And use the following formula:

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

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?

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

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:

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

For (Income, Class=No):

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

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

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

Post-sessional work

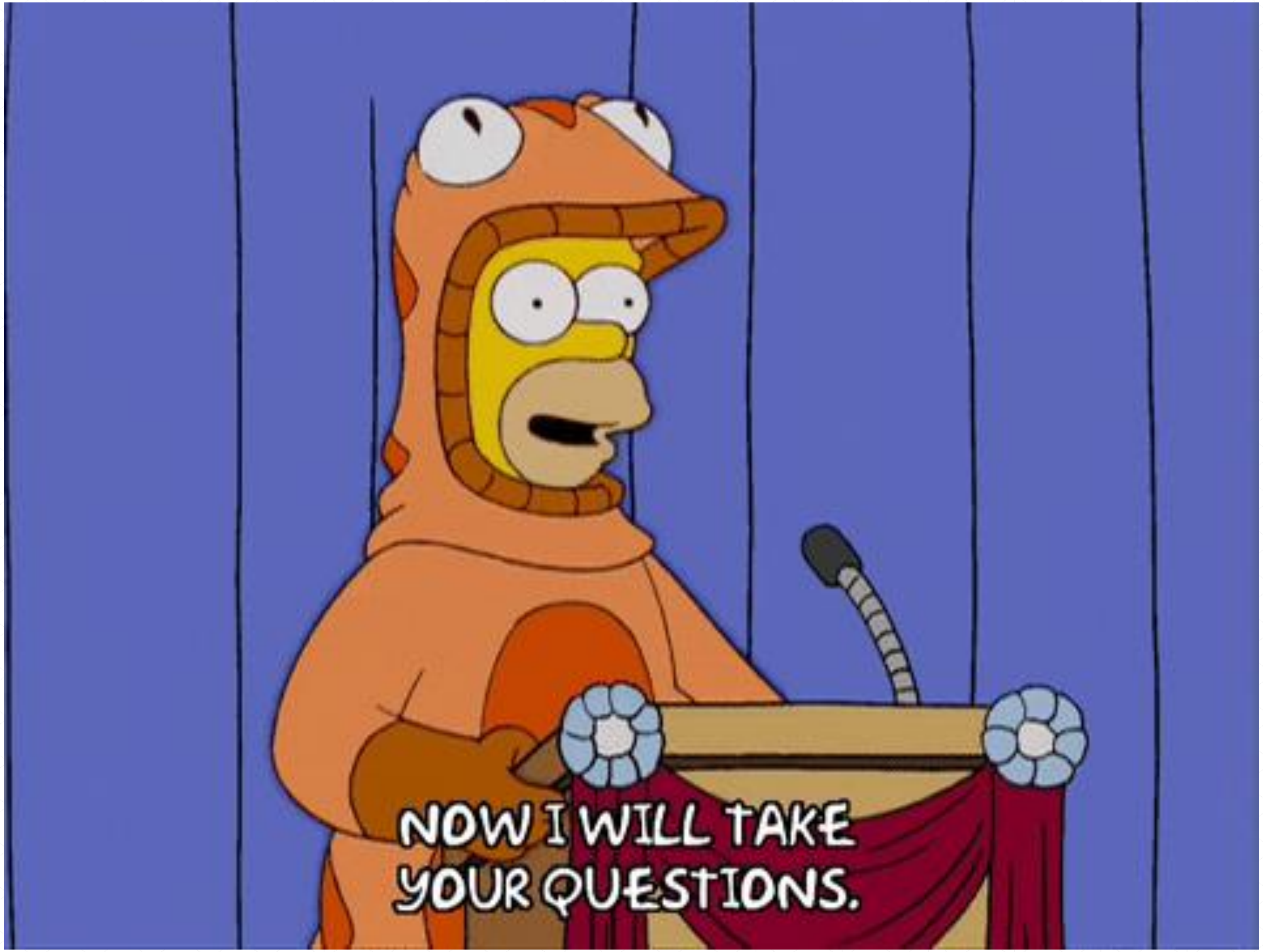
- Write a report on what is a Bayesian Network and how it can be used in real-world problems. Make sure to add references to your report.

References

- Callan, R. (2003). *Artificial Intelligence*. Part-3. Palgrave Macmillan

Next Session!

- Probability Part 3


NOW I WILL TAKE
YOUR QUESTIONS.