Bayes Theorem

Probability

- Probability is the measure of the likelihood that an event will occur. Probability is quantified as a number between 0 and 1 (where 0 indicates impossibility and 1 indicates certainty).

- Example A simple example is the toss of a fair (unbiased) coin. Since the two outcomes are equally probable, the probability of "heads" equals the probability of "tails", so the probability is 1/2 (or 50%) chance of either "heads" or "tails".

Probability

- **Marginal Probability**: The probability of an event irrespective of the outcomes of other random variables, e.g. P(A).
- **Joint Probability**: Probability of two (or more) simultaneous events, e.g. P(A and B) or P(A, B).

Two dependent events

$$P(A, B) = P(A \mid B) * P(B)$$

- **Conditional Probability**: Probability of one (or more) event given the occurrence of another event, e.g. P(A given B) or P(A | B).

  - $P(A \mid B) = P(A \cap B) / P(B)$

Bayes Theorem

Bayes Theorem: Principled way of calculating a conditional probability without the joint probability.

The result P(A|B) is referred to as the **posterior probability** and P(A) is referred to as the **prior probability**.

- P(A|B): Posterior probability.
- P(A): Prior probability.

P(B|A) is referred to as the **likelihood** and P(B) is referred to as the **evidence**.

- P(B|A): Likelihood.
- P(B): Evidence.
- Posterior = Likelihood * Prior / Evidence

Bayes Theorem

Bayes' rule
or
Bayes' theorem

LIKELIHOOD
The probability of "B" being True, given "A" is True

PRIOR
The probability "A" being True. This is the knowledge.

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

POSTERIOR
The probability of "A" being True, given "B" is True

MARGINALIZATION
The probability "B" being True.

Bayes Theorem

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

Bayes Theorem

Where A and B are events:

- P(A) and P(B) are the probabilities of A and B without regard to each other.
- P(A | B), a conditional probability, is the probability of observing event A given that B is true.
- P(B | A) is the probability of observing event B given that A is true.

Conditional probability

- Conditional probability is a measure of the likelihood of an event occurring provided that another event has already occurred (through assumption, supposition, statement, or evidence).

- If A is the event of interest and B is known or considered to have occurred, the conditional probability of A given B is generally stated as $P(A|B)$ or, less frequently, $P_B(A)$ if A is the event of interest and B is known or thought to have occurred.

Bayesian Rule

- ***Class prior or prior probability:*** probability of event A occurring before knowing anything about event B.
- ***Predictor prior or evidence:*** same as class prior but for event B.
- ***Posterior probability:*** probability of event A after learning about event B.
- ***Likelihood:*** reverse of the posterior probability.

Bayes Theorem - Special Case

$$P(H|E) = \frac{P(E|H)\,P(H)}{P(E)}$$

Here, H is Hypothesis, E is Evidence, '|' is 'given that'.

$$P(H|E) = \frac{P(E|H)\,P(H)}{P(E|H)P(H) + P(E|\neg H)P(\neg H)}$$

Example

Marie will have a birthday party tomorrow, in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Marie's party?

$P\left(\begin{smallmatrix}\text{correct}\\ \text{predict}\end{smallmatrix}\right) = ?$

Ans) $P(\text{Rain}) = \frac{5}{365}$

$P(\text{Not Rain}) = \frac{360}{365}$

$P\left(\begin{smallmatrix}\text{correct}\\ \text{Predict}\end{smallmatrix} \middle| \text{Rain}\right) = \frac{90}{100}$

$\therefore P\left(\begin{smallmatrix}\text{Not}\\ \text{correct Predict}\end{smallmatrix} \middle| \text{Rain}\right) = \frac{10}{100}$

$P\left(\begin{smallmatrix}\text{Not}\\ \text{correct predict}\end{smallmatrix} \middle| \begin{smallmatrix}\text{Not}\\ \text{Rain}\end{smallmatrix}\right) = \frac{10}{100}$

$\therefore P\left(\text{correct predict} \middle| \begin{smallmatrix}\text{Not}\\ \text{Rain}\end{smallmatrix}\right) = \frac{90}{100}$

$P\left(\text{Rain} \middle| \begin{smallmatrix}\text{correct}\\ \text{predict}\end{smallmatrix}\right) = ?$

Hence,

$$P\left[\text{Rain} \middle| \begin{smallmatrix}\text{correct}\\ \text{predict}\end{smallmatrix}\right] = \frac{P\left(\begin{smallmatrix}\text{correct}\\ \text{predict}\end{smallmatrix} \middle| \text{Rain}\right) P(\text{Rain})}{P\left(\begin{smallmatrix}\text{correct}\\ \text{pred.}\end{smallmatrix} \middle| \text{Rain}\right) P(\text{Rain}) + P\left(\begin{smallmatrix}\text{correct}\\ \text{pred.}\end{smallmatrix} \middle| \begin{smallmatrix}\text{Not}\\ \text{Rain}\end{smallmatrix}\right) P\left(\begin{smallmatrix}\text{Not}\\ \text{Rain}\end{smallmatrix}\right)}$$

$$= \frac{\left(\frac{90}{100}\right)\left(\frac{5}{365}\right)}{\left(\frac{90}{100}\right)\left(\frac{5}{365}\right) + \left(\frac{90}{100}\right)\left(\frac{350}{365}\right)}$$

$\simeq 0.014$

$P(R|CP) = ?$

$P(R) = \frac{5}{365}$

$P(NR) = \frac{360}{365}$

$P(CP|R) = \frac{90}{100}$

$P(NCP|R) = \frac{10}{100}$

$P(NCP|NR) = \frac{10}{100}$

$P(CP|NR) = \frac{90}{100}$

Bayes Interference

Bayesian inference Bayesian inference is a method of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as evidence. It Involves:

Prior Probability: The initial Probability based on the present level of information.

Posterior Probability: A revised Probability based on additional information.

Solution

The sample space is defined by two events - it rains or it does not rain. Additionally, a third event occurs when the weatherman predicts rain. Notation for these events appears below

Event A1. It rains on Marie's party.
Event A2. It does not rain on Marie's party.

Event B. The weatherman predicts rain.

Solution

In terms of probabilities, we know the following:

P(A1) = 5/365 =0.0136985 [It rains 5 days out of the year.]
P(A2) = 360/365 = 0.9863014 [It does not rain 360 days out of the year.]
P(B | A1) = 0.9 [When it rains, the weatherman predicts rain 90% of the time.]
P(B | A2) = 0.1 [When it does not rain, the weatherman predicts rain 10% of the time.]

Solution

- $P(A_1) = 5/365 = 0.0136985$ [It rains 5 days out of the year.]
- $P(A_2) = 360/365 = 0.9863014$ [It does not rain 360 days out of the year.]
- $P(B \mid A_1) = 0.9$ [When it rains, the weatherman predicts rain 90% of the time.]
- $P(B \mid A_2) = 0.1$ [When it does not rain, the weatherman predicts rain 10% of the time.]

Solution

$$P(A_1 \mid B) = \frac{P(A_1)\,P(B \mid A_1)}{P(A_1)\,P(B \mid A_1) + P(A_2)\,P(B \mid A_2)}$$

$$P(A_1 \mid B) = (0.014)(0.9) \,/\, [\,(0.014)(0.9) + (0.986)(0.1)\,]$$

$$P(A_1 \mid B) = 0.111$$

A Bayesian Network

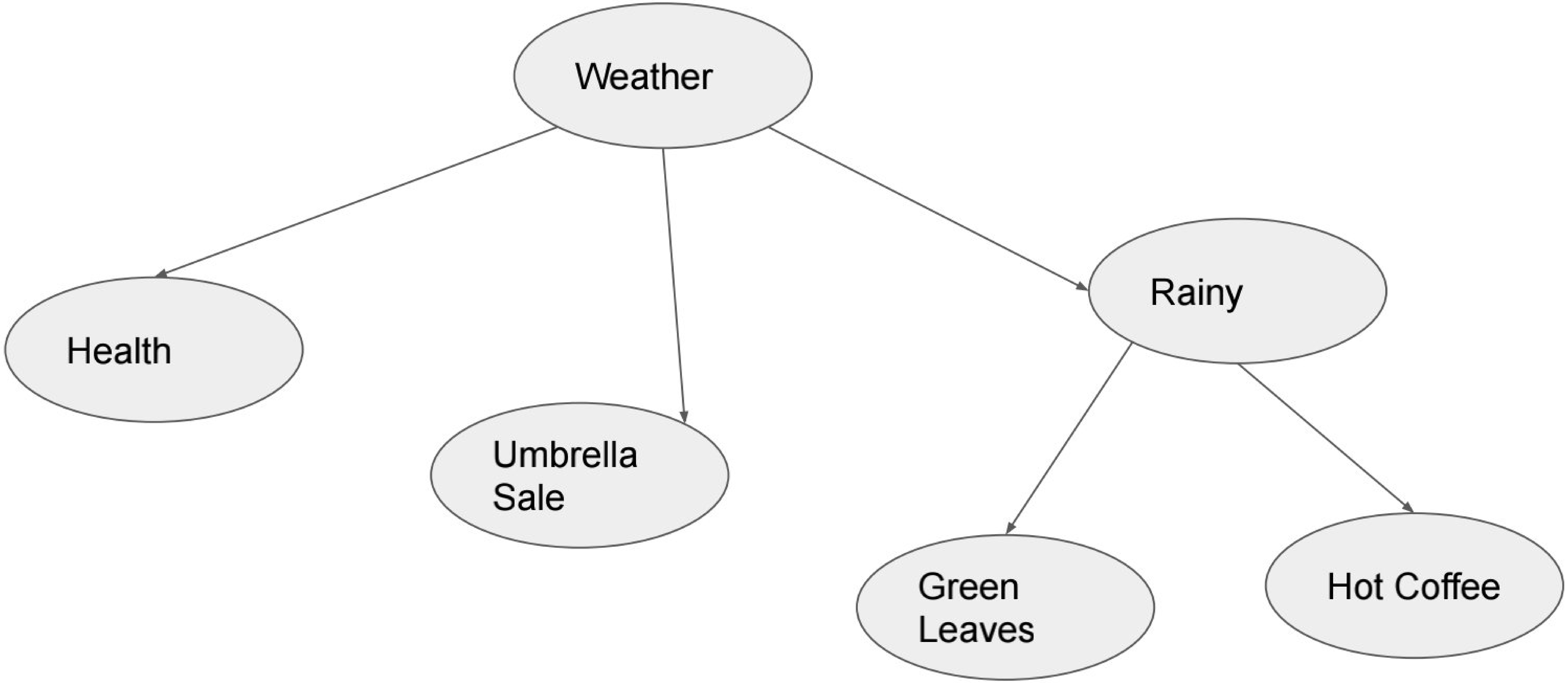
A Bayesian network (also known as a Bayes network, belief network, or decision network) is a probabilistic graphical model or graph data structure.

Each node represents a random variable and its conditional dependencies via a directed acyclic graph (DAG).

Bayesian Belief Network

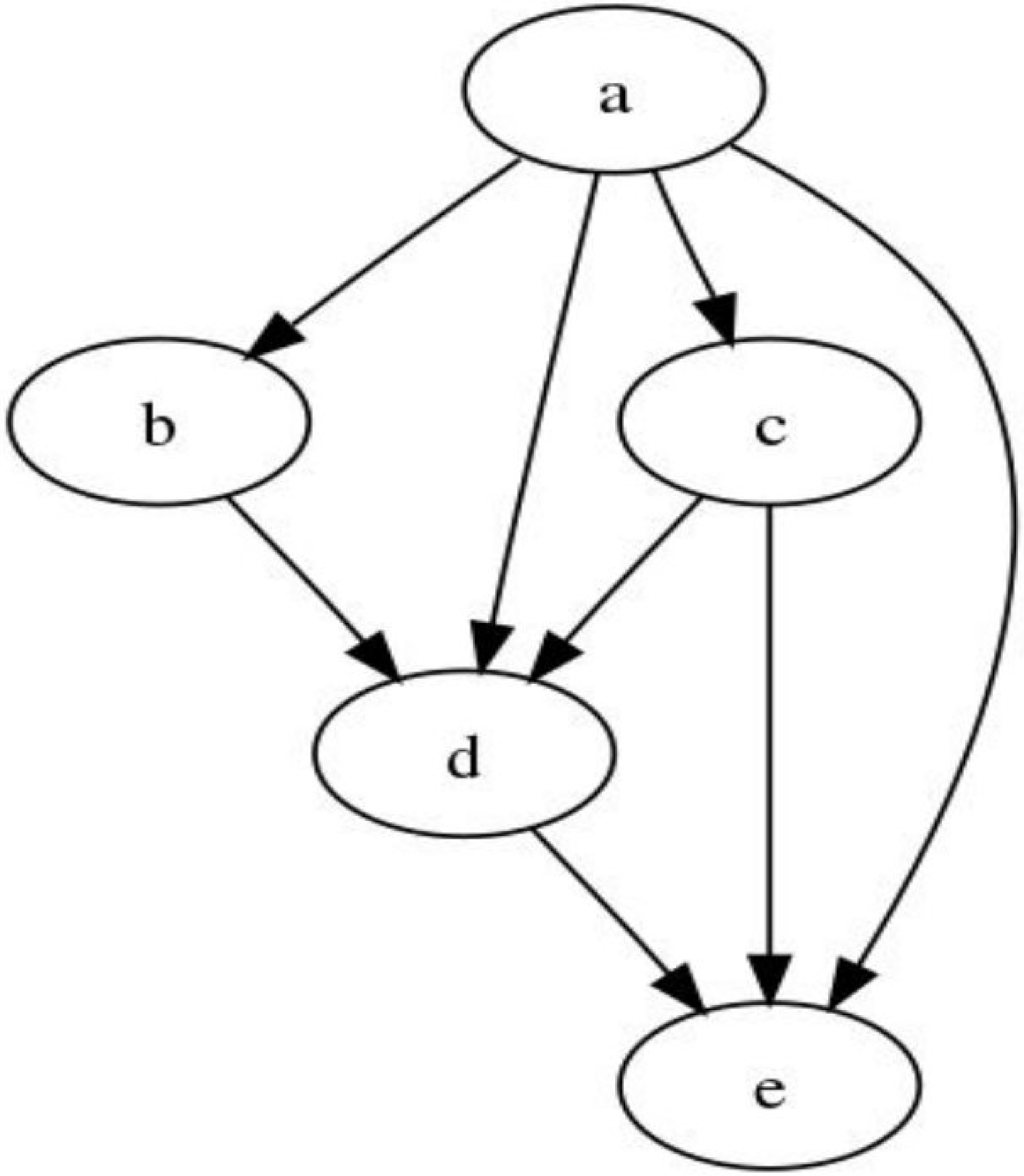
- Other names: Bayes network, belief network, decision network, or Bayesian model.
- A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph
- The graph consists of graph nodes and arcs.
- The node represents variables which can be discrete or continuous.
- The arc represent casual relationships.

Example



Weather
Health
Umbrella Sale
Rainy
Green Leaves
Hot Coffee

Directed Acyclic Graph



Directed Acyclic Graph

- In graph theory and computer science, a directed acyclic graph (DAG) is a directed graph with no directed cycles.

- In other words, it’s made up of vertices and edges (also called arcs), with each edge pointing from one vertex to the next in such a way that following those directions would never lead to a closed-loop as depicted in below picture.

Ch – NaiveBayes, Bayes theorem, Bayesian Network

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

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

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

| Slide ⇒ 11, Math | CT 2 ⇒ Math |

Q) Why Belief Network or Bayes Network is a DAG?

→ Belief network

A graphical model that represents a set of variables with nodes and their probabilistic conditional dependencies with direct arcs.

→ DAG is a direct graph with each edge pointing to vertices in such way that no closed loop occurs.

→ Now, if a Bayes network is cyclic (unlike a DAG), then final decision making will never occur.

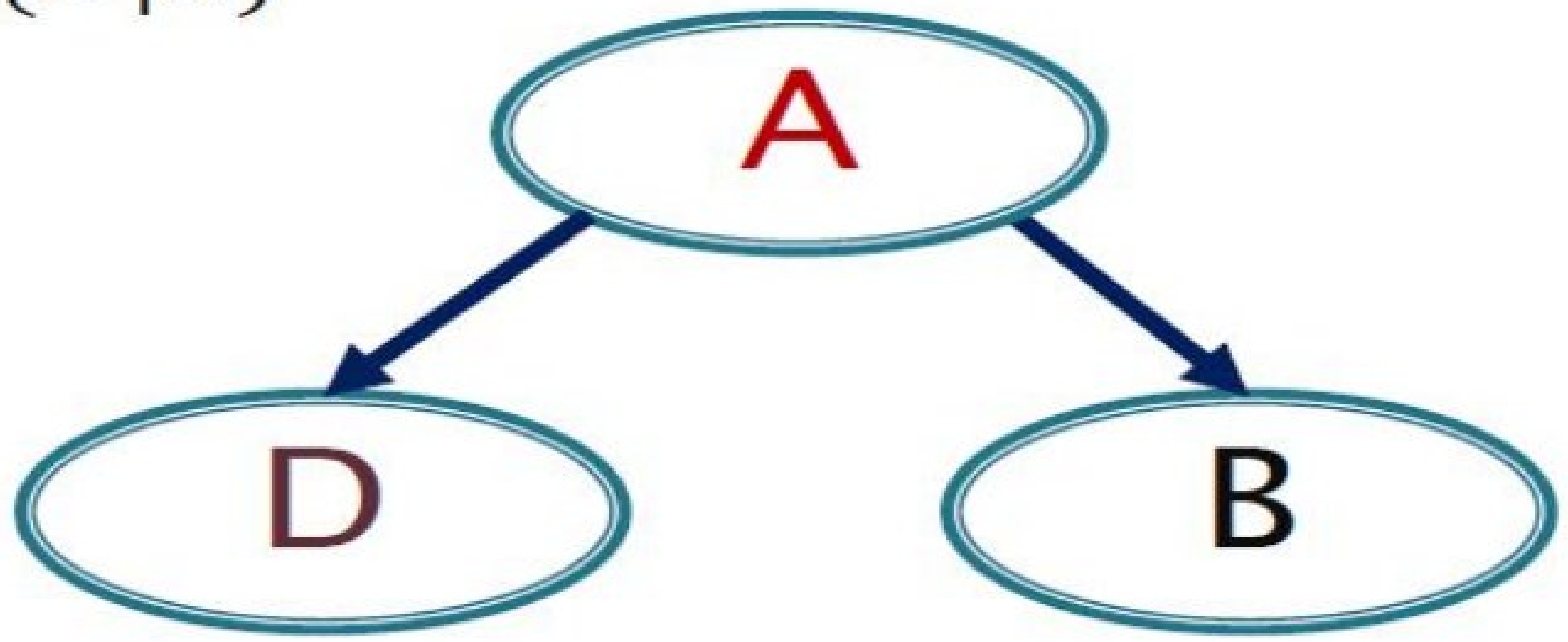
Bayesian Belief Network

$$P(X_i|X_{i-1}, \ldots\ldots, X_1) = P(X_i \,|Parents(X_i\,))$$

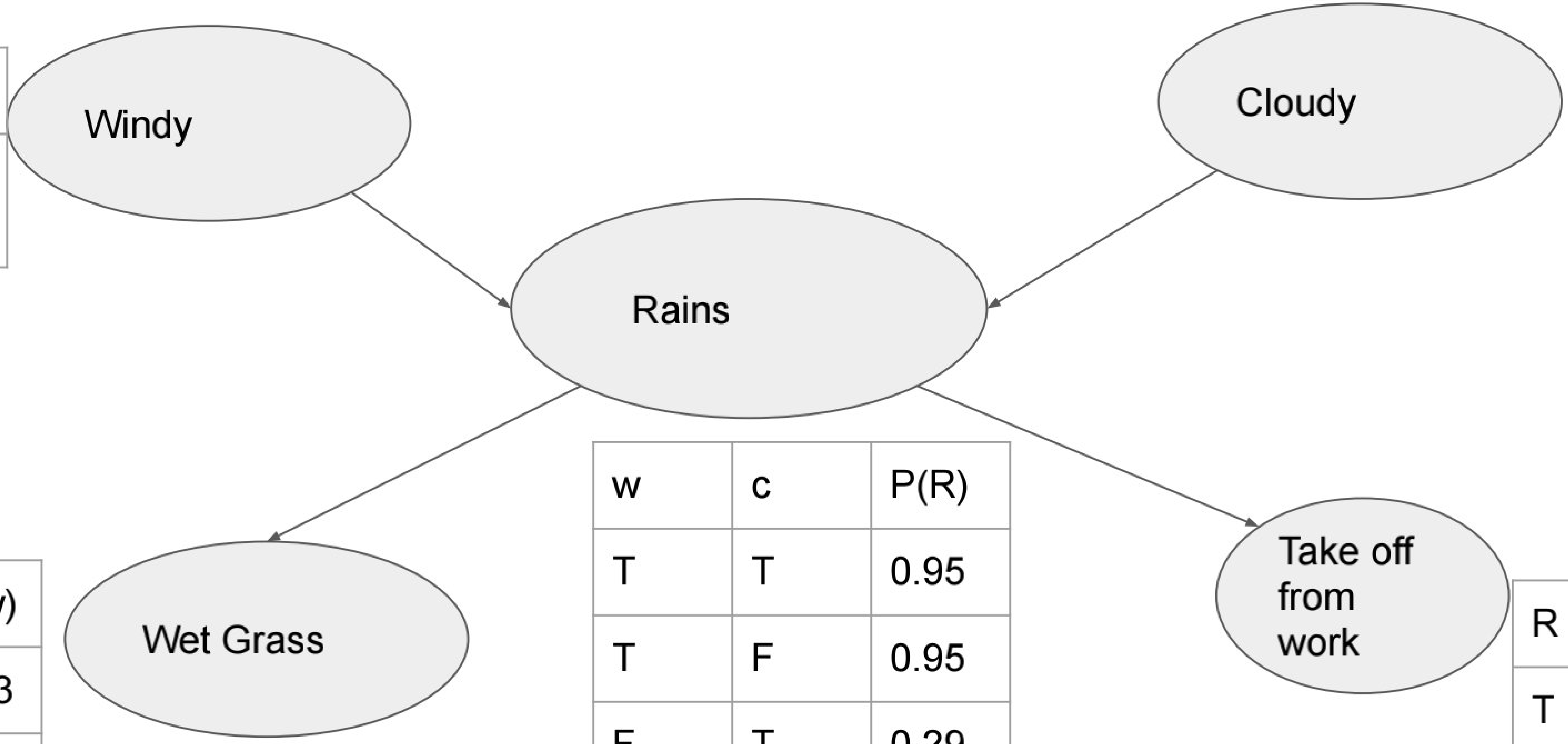
$$P(x_1, x_2, x_3, \ldots \ldots, x_n) = \prod_{i=1}^{n} P(x_i|Parent(x_i))$$

So,

$$P(A, D, B) \equiv P(A) \times P(B|A) \times P(D|A)$$



Example



P(w)
0.001

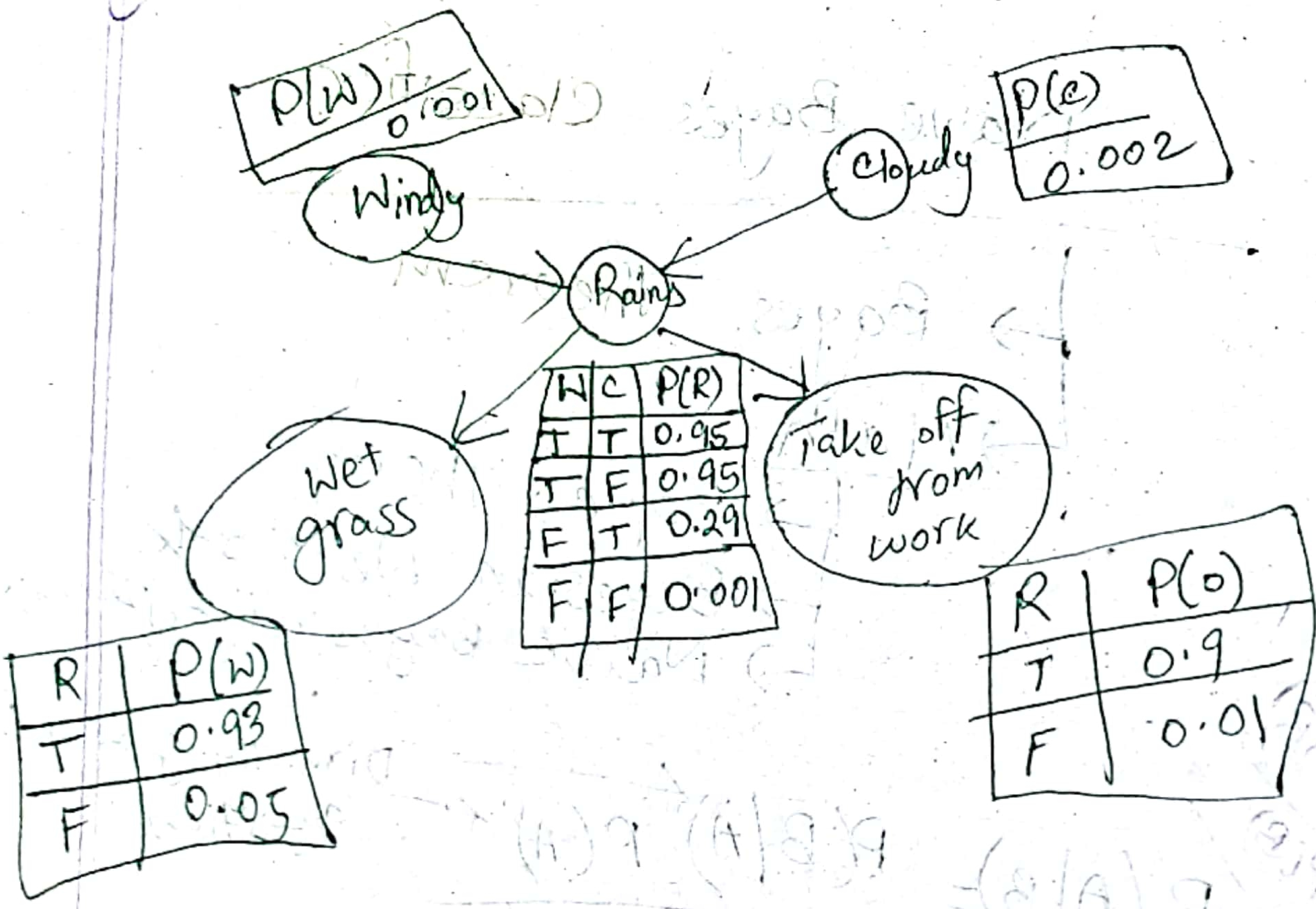
P(c)
0.002

w	c	P(R)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

R	P(w)
T	0.93
F	0.05

R	P(o)
T	0.9
F	0.01

2)



P(W)
0.001

P(C)
0.002

W	C	P(R)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

R	P(W)
T	0.93
F	0.05

R	P(O)
T	0.9
F	0.01

$P(W)_{\text{Wet grass}} = ?$

$$P(W) = P(W/R) \times P(R) + P(W/\bar{R}) \times P(\bar{R})$$

$$= (0.93 \times P(R)) + (P(\bar{R}) \times 0.05)$$

eq (1)

$$P(R) = P(R \mid Wd, cl) \times P(Wd \wedge cl)$$

$$+$$

$$P(R \mid Wd, \overline{cl}) \times P(Wd \wedge \overline{cl})$$

$$+$$

$$P(R \mid \overline{Wd}, \overline{cl}) \times P(\overline{Wd} \wedge \overline{cl})$$

$$+$$

$$P(R \mid \overline{Wd}, cl) \times P(\overline{Wd} \wedge cl)$$

$$= \left(0.95 \times (0.001 \times 0.002)\right) + \left(0.95 \times (0.001 \times (1-0.002))\right)$$

$$+ \left(0.29 \times ((1-0.001) \times 0.002)\right) + \left(0.001 \times ((1-0.001) \times (1-0.002))\right)$$

$$= \ldots$$

$$P(\overline{R}) = 1 - P(R) = \ldots$$

Substitute in eq 1 to find

$P(\text{wetgrass})$

or $P(W)$

Example

Find the probability of having wet grass.

Naive Bayes Classifier

- Naive Bayes is a set of simple and efficient machine learning algorithms for solving a variety of classification and regression problems.
- Naive Bayes Classifiers are based on the Bayes Theorem.
- One assumption taken is the strong independence assumptions between the features. These classifiers assume that the value of a particular feature is independent of the value of any other feature. And so they are referred as being "Naive".
- In a supervised learning situation, Naive Bayes Classifiers are trained very efficiently.
- Naive Bayes classifiers need a small training data to estimate the parameters needed for classification.
- Naive Bayes Classifiers have simple design and implementation and they can applied to many real life situations.

Naive Bayes Classification

Bayes Theorem $\longrightarrow$ $$P(A \mid x_1, \ldots, x_n) = \frac{P(x_1, \ldots, x_n \mid A)\, P(A)}{P(x_1, \ldots, x_n) \Rightarrow 1}$$

since features are independent

$\downarrow$

Naive Bayes $\longrightarrow$ $$P(A \mid x_1, \ldots, x_n) = P(x_1 \mid A) \cdot P(x_2 \mid A) \cdot P(x_i \mid A)\, P(A)$$

Naive Bayes Classification

The Naive Bayes classifier makes two fundamental assumptions on the observations.

- The target classes are **independent** of each other. Consider a rainy day with strong winds and high humidity. A Naive classifier would treat these two features, wind and humidity, as independent parameters. That is to say, each feature would impose its probabilities on the outcome, such as rain in this case.
- Prior probabilities for the target classes are **equal**. That is, before calculating the posterior probability of each class, the classifier will assign each target class the same prior probability.

Gaussian Naive Bayes

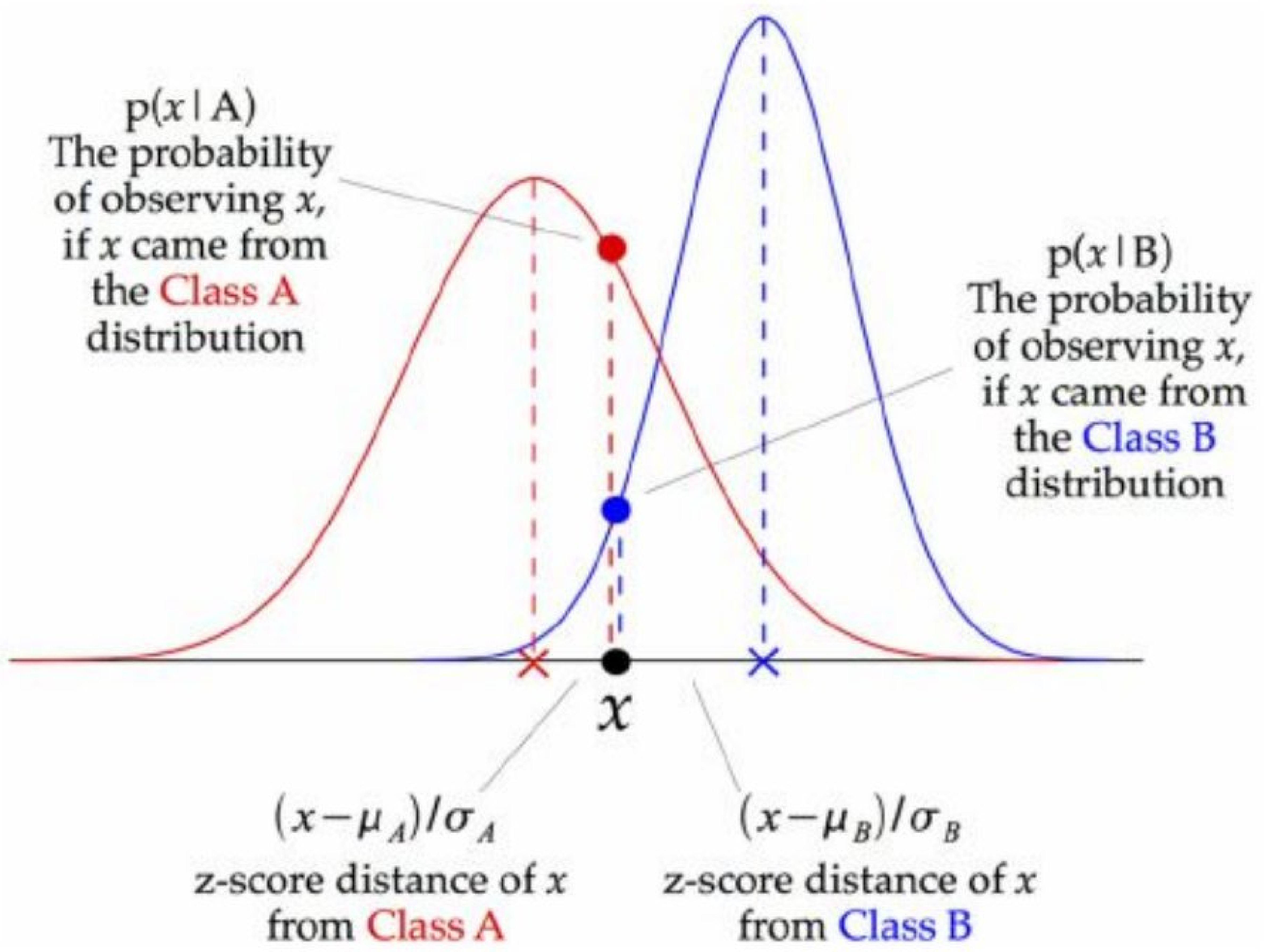
- **Gaussian Naive Bayes** is a variant of Naive Bayes that follows Gaussian normal distribution and supports continuous data. We have explored the idea behind Gaussian Naive Bayes along with an example.
- When working with continuous data, an assumption often taken is that the continuous values associated with each class are distributed according to a normal (or Gaussian) distribution. The likelihood of the features is assumed to be-

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

Gaussian Naive Bayes

- Gaussian Naive Bayes supports continuous valued features and models each as conforming to a Gaussian (normal) distribution.
- An approach to create a simple model is to assume that the data is described by a Gaussian distribution with no co-variance (independent dimensions) between dimensions.
- This model can be fit by simply finding the mean and standard deviation of the points within each label, which is all what is needed to define such a distribution.

Gaussian Naive Bayes



Types of Naive Bayes

- The **Multinomial Naive Bayes** method is a common Bayesian learning approach in natural language processing. Using the Bayes theorem, the program estimates the tag of a text, such as an email or a newspaper piece. It assesses the likelihood of each tag of multinomial Naive Bayes for a given sample and returns the tag with the highest possibility.
- The **Bernoulli Naive Bayes** is a part of the family of Naive Bayes. It only takes binary values. Multiple features may exist, but each is assumed to be a binary-valued (Bernoulli, boolean) variable. Therefore, this class requires samples to be represented as binary-valued feature vectors.
- The **Gaussian Naive Bayes** is a variant of Naive Bayes that follows Gaussian normal distribution and supports continuous data. To build a simple model using Gaussian Naive Bayes, we assume the data is characterized by a Gaussian distribution with no covariance (independent dimensions) between the parameters. This model may fit by applying the Bayes theorem to calculate the mean and standard deviation of the points within each label.

Bernoulli Naive Bayes

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	Yes	Pass

Example

Outlook	Temperature	Humidity	Windy	Play?
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Example

outlook			temperature			humidity			windy			play?	
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	2	3	hot	2	2	high	3	4	FALSE	6	2	9	5
overcast	4	0	mild	4	2	normal	6	1	TRUE	3	3		
rainy	3	2	cool	3	1								

Example

outlook			temperature			humidity			windy			play?	
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	2/9	3/5	hot	2/9	2/5	high	3/9	4/5	FALSE	6/9	2/5	9/14	5/14
overcast	4/9	0/5	mild	4/9	2/5	normal	6/9	1/5	TRUE	3/9	3/5		
rainy	3/9	2/5	cool	3/9	1/5								

Example

X = {outlook: sunny, temperature: mild, humidity: normal, windy: false}.

$$P(yes) = 9/14$$

$$P(outlook = sunny|yes) = 2/9$$

$$P(temperature = mild|yes) = 4/9$$

$$P(humidity = normal|yes) = 6/9$$

$$P(windy = false|yes) = 6/9$$

Now find if Play yes or No when

$X = \{$outlook: sunny, temperature: mild, humidity: normal, windy: False$\}$

Ans) $P(\text{yes}) = \frac{8}{14}$ $\qquad P(\text{No}) = \frac{6}{14}$

$P(\text{outlook: sunny}/\text{yes}) = 2/8$

$* \; P(\text{temp: mild}/\text{yes}) = 3/8$

$* \; P(\text{wind: False}/\text{yes}) = 3/8$

$* \; P(\text{humidity: normal}/\text{Yes}) = 5/8$

$\frac{2}{8} \times \frac{3}{8} \times \frac{3}{8} \times \frac{5}{8} = 0.02197$

$\times \frac{8}{14}$

Similarly,

$P(\text{sunny}/\text{No}) *$
$P(\text{mild}/\text{No}) *$
$P(\text{False}/\text{No}) *$
$P(\text{normal}/\text{No})$

$= \frac{3}{6} \times \frac{1}{6} \times \frac{5}{6} \times \frac{4}{6}$
$\times \frac{6}{14}$
$= 0.04630$

$\Downarrow$

So, Play No

Because $P(\text{Play yes}) < P(\text{Play No})$

শুধুমাত্র probability enough না for achieving better accuracy,

↳ So we use decision tree which deals with feature importance, entropy.

Example

X = {outlook: sunny, temperature: mild, humidity: normal, windy: false}.

$$P(yes|X) \propto P(X|y) * P(y)$$

$$P(yes|X) \propto P(x_1|y) * P(x_2|y) * P(x_3|y) * P(x_4|y) * P(y)$$

$$P(yes|X) \propto P(sunny|yes) * P(mild|yes) * P(normal|yes) * P(false|yes) * P(yes)$$

$$P(yes|X) \propto \frac{2}{9} * \frac{4}{9} * \frac{6}{9} * \frac{6}{9} * \frac{9}{14}$$

$$\boldsymbol{P(yes|X) \propto 0.0282}$$

$$\boldsymbol{P(no|X) \propto 0.0069}$$