



**BITS Pilani**  
Pilani | Dubai | Goa | Hyderabad

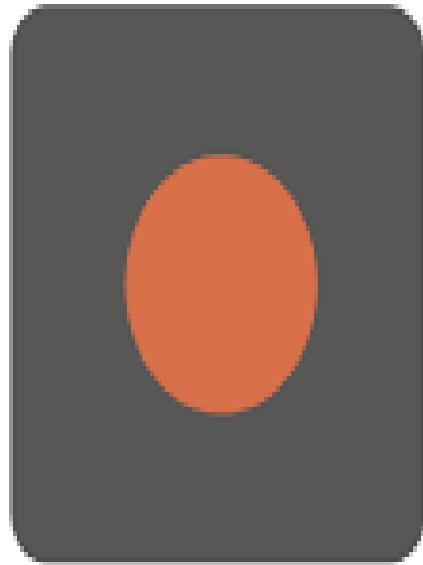
# Introduction to Statistical Methods

---

**ISM Team**

# IMP Note to Self

---



Start  
Recording

# RECAPITULATION

---

- **Addition Rule of Probability**
- **Conditional Probability**
- **Independent events**
- **Total Probability**



**Baye's theorem and Naïve Bayes theorem**  
**(Session 4: 14<sup>th</sup>&15<sup>th</sup> December 2024)**

Contact Session	List of Topic Title	Reference
CS - 4	Bayes theorem(with proof),Introduction to Naïve Bayes concept.	T1 & T2
HW	Problems on Bayes theorem	T1 & T2

## BAYES' THEOREM:

---

Suppose that  $E_1, E_2, \dots, E_n$  are mutually exclusive events of a sample space “S” such that  $P(E_i) > 0$  for  $i = 1, 2, 3, \dots, n$  and A is any arbitrary event of “S”

such that  $P(A) > 0$  and  $A \subseteq \bigcup_{i=1}^n E_i$  then the conditional probability of  $E_i$  given A

$$\text{for } i = 1, 2, 3, \dots, n \text{ is given by } P(E_i | A) = \frac{P(E_i)P(A|E_i)}{\sum_i^n P(E_i)P(A|E_i)}$$

## Proof:

Given that  $A \subseteq \bigcup_{i=1}^n E_i$

$$\therefore A = A \cap \left( \bigcup_{i=1}^n E_i \right)$$

$$\Rightarrow A = A \cap (E_1 \cup E_2 \cup \dots \cup E_n)$$

$$A = (A \cap E_1) \cup (A \cap E_2) \cup \dots \cup (A \cap E_n) \quad (\text{Distributive law})$$

$$\text{ie } A = \bigcup_{i=1}^n (A \cap E_i)$$

Since  $E_1, E_2, \dots, E_n$  are mutually exclusive events,  $E_i \cap E_j = \emptyset$  for  $i \neq j$

and  $A \cap E_1, A \cap E_2, \dots, A \cap E_n$  are also mutually exclusive

$$(A \cap E_i) \cap (A \cap E_j) = \emptyset \quad \text{for } i \neq j$$

Applying the addition rule for mutually exclusive events, we have

$$\begin{aligned} P(A) &= P((A \cap E_1) \cup (A \cap E_2) \dots \cup (A \cap E_n)) \\ &= P(A \cap E_1) + P(A \cap E_2) \dots \dots + P(A \cap E_n) \end{aligned}$$

$$P(A) = P\left(\bigcup_{i=1}^n (A \cap E_i)\right) = \sum_{i=1}^n P(A \cap E_i) \quad \text{--- (1)}$$

From multiplication rule we have

$$P(A \cap E_i) = P(E_i) P(A/E_i) \quad \text{----- (2)} \quad \text{Substituting (2) in (1) we have}$$

$$P(A) = \sum_{i=1}^n P(E_i) P(A/E_i) \quad \text{--- (3)[Rule of total probability OR Rule of elimination ]}$$

---

From the definition of conditional probability

$$P(A / E_i) = \frac{P(E_i \cap A)}{P(E_i)} \quad \dots \quad (4)$$

Substituting (2) and (3) in (4) we get

$$P(E_i / A) = \frac{\sum_i P(E_i)P(A/E_i)}{\sum_i P(E_i)}$$

## Example:1

---

The Probabilities of  $X, Y$  and  $Z$  becoming managers are  $\frac{4}{9}, \frac{2}{9}$ , and  $\frac{1}{3}$  respectively.

The probabilities that the bonus schemes will be introduced if  $X, Y$  and  $Z$  becomes managers are  $\frac{3}{10}, \frac{1}{2}$  and  $\frac{4}{5}$  respectively.

- (i) What is the probability that bonus scheme will be introduced ?
- (ii) If the bonus scheme has been introduced, What is the probability that the manager appointed was  $X$ ?

## Solution:

Let  $E_1, E_2, E_3$  denote the events that X, Y and Z become Managers respectively and A denote the event that Bonus scheme is introduced.

$$P(E_1) = \frac{4}{9}, P(E_2) = \frac{2}{9}, P(E_3) = \frac{1}{3}, P(A|E_1) = \frac{3}{10}, P(A|E_2) = \frac{1}{2}, P(A|E_3) = \frac{4}{5}$$

(i) Using Addition theorem of Probability

$$\begin{aligned} P(A) &= P(E_1 \cap A) + P(E_2 \cap A) + P(E_3 \cap A) \\ &= P(E_1)P(A|E_1) + P(E_2)P(A|E_2) + P(E_3)P(A|E_3) = \frac{4}{9} \times \frac{3}{10} + \frac{2}{9} \times \frac{1}{2} + \frac{1}{3} \times \frac{4}{5} = \frac{23}{45} \end{aligned}$$

(ii) Using Baye's Theorem,

$$P(E_1|A) = \frac{P(E_1)P(A|E_1)}{P(A)} = \frac{6}{23}$$

## Example: 2

In a neighborhood, 90% children were falling sick due to flu and 10% due to measles and no other disease. The probability of observing rashes for measles is 0.95 and for flu is 0.08. If a child develops rashes, Find the child's probability of having flu.

### Solution:

Let, F: children with flu M: children with measles

R: children showing the symptom of rash

$$P(F) = 90\% = 0.9, P(M) = 10\% = 0.1$$

$$P(R|F) = 0.08, P(R|M) = 0.95$$

$$P(F|R) = \frac{P(R|F)P(F)}{P(R|M)P(M)+P(R|F)P(F)}$$

$$P(F|R) = \frac{0.08 \times 0.9}{0.95 \times 0.1 + 0.08 \times 0.9} = 0.43$$

## Example:3

---

It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam e-mails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

## Solution:

Let  $A$  = event that an email is detected as spam,  $B$  = event that an email is spam,  
 $B^c$  = event that an email is not spam.

We know  $P(B) = P(B^c) = .5$ ,

$P(A | B) = 0.99$ ,  $P(A | B^c) = 0.05$ .

Hence by the Bayes's formula

$$\text{we have } P(B^c | A) = \frac{P(A | B^c)P(B^c)}{P(A | B)P(B) + P(A | B^c)P(B^c)}$$

$$= (0.05 \times 0.5) / (0.05 \times 0.5 + 0.99 \times 0.5) = \frac{5}{104}$$

## Example :4

If the weather is sunny, Then the player will play or not?  
i.e., Play /sunny = Yes or No.

Note: If we know  $P(\text{Yes/Sunny})$  and  $P(\text{No/Sunny})$   
then we can answer the question asked

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

# Steps to Apply Bayes' Theorem

Step1: View or Collect “raw” data

Step2: Convert long data into frequency table

Weather	No	Yes
Sunny	2	3
Overcast	0	4
Rainy	3	2

Step3: Row and Column sums to get Probabilities

Weather Probabilities: Sunny =5/14, Rainy=5/14, Overcast= 4/14

Play Probabilities :

No =5/14

Yes =9/14

Weather	No	Yes	Row Total	
				P(sunny)= 5/14
Sunny	2	3	5	P(sunny)= 5/14
Overcast	0	4	4	P(overcast)=4/14
Rainy	3	2	5	P(Rainy)= 5/14
Column Total	5	9	14	
P(no)=5/14 P(yes)=9/14				

Step4: Apply Probabilities from frequency table to Bayes Theorem

$$P(\text{ Yes} \mid \text{Sunny}) = \frac{P(\text{Sunny}|\text{yes})P(\text{yes})}{P(\text{Sunny})} = \frac{\frac{3}{9} \times \frac{9}{14}}{\frac{5}{14}} = 0.60, \quad P(\text{ No} \mid \text{Sunny}) = \frac{P(\text{Sunny}|\text{No})P(\text{No})}{P(\text{Sunny})} = \frac{\frac{2}{5} \times \frac{5}{14}}{\frac{5}{14}} = 0.40$$

$P(\text{ Yes} \mid \text{Sunny}) > P(\text{ No} \mid \text{Sunny})$ . Hence, The Player will play the game.

# Bayes' Theorem

---

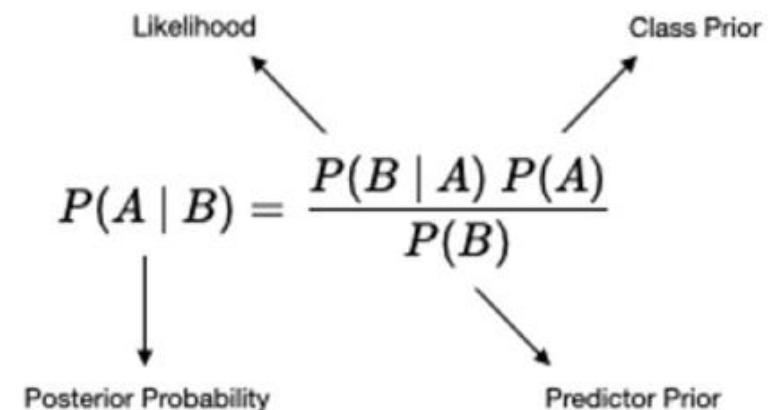
- $P(h)$  = prior probability of hypothesis  $h$ , before seeing the training data
- $P(D)$  = prior probability of training data  $D$
- $P(h | D)$  = probability of  $h$  given  $D$
- $P(D | h)$  = probability of  $D$  given  $h$

$$P(h | D) = \frac{P\left(\frac{D}{h}\right)P(h)}{P(D)}$$

# Generative Model

- ❖ Build model to estimate the posterior probability  $P(Y | X)$  by estimating
- ❖ likelihood of data given target (hypothesis)  $P(X | Y)$
- ❖ Prior probabilities over target  $P(Y)$
- ❖ In general, for a specific class  $Y=c_k$ ,

$$P(Y = c_k | X) = \frac{P(X|Y = c_k) * P(Y=c_k)}{P(X)}$$



# Choosing hypothesis

Generally want the most probable hypothesis given the training data

Maximum a posteriori hypothesis  $h_{MAP}$ :

$$\begin{aligned}
 h_{MAP} &= \arg \max_{h \in H} P(h|D) \\
 &= \arg \max_{h \in H} \frac{P(D|h).P(h)}{P(D)} \\
 &= \arg \max_{h \in H} P(D|h).P(h)
 \end{aligned}$$

If assume  $P(h_i) = P(h_j)$  then can further simplify, and choose the Maximum likelihood (ML) hypothesis

$$h_{ML} = \arg \max_{h_i \in H} P(D|h_i)$$

# Bayesian Learning

---

- Naive Bayes is a set of simple and efficient machine learning algorithms for solving a variety of classification and regression problems.
- Naive Bayes assumes conditional independence where Bayes theorem does not. This means the relationship between all input features are independent.
- Bayesian learning algorithms that calculate explicit probabilities for hypotheses, such as the Naive Bayes classifier, are among the most practical approaches to certain types of learning problems.

Example: Problem of learning to classify text documents such as electronic news articles.

- For such learning tasks, the Naive Bayes classifier is among the most effective known algorithms.

## Features of Bayesian learning

---

- Prior knowledge can be combined with observed data to determine the final probability of a hypothesis.
- Prior knowledge is provided by asserting
  - ❖ prior probability for each candidate hypothesis, and
  - ❖ probability distribution over observed data for each possible hypothesis.
- New instances can be classified by combining the predictions of multiple hypothesis, weighted by their probabilities.

# Conditional Independence

---

**Definition:**  $X$  is conditionally independent of  $Y$  given  $Z$ , if the probability distribution governing  $X$  is independent of the value of  $Y$ , given the value of  $Z$

$$(\forall i, j, k) P(X=x_i | Y = y_j, Z = z_k) = P(X=x_i | Z_k)$$
$$P(X|Y, Z) = P(X | Z)$$

**Example:**

$$P(\text{Thunder} | \text{Rain, Lightning}) = P(\text{Thunder} | \text{Lightning})$$

## Applying Conditional Independence

---

Naive Bayes assumes  $X_i$  are conditionally independent given Y.

e.g.,  $P(X_1|X_2, Y) = P(X_1|Y)$

$$\begin{aligned}P(X_1, X_2|Y) &= P(X_1|X_2, Y) P(X_2|Y) \\&= P(X_1|Y) P(X_2|Y)\end{aligned}$$

General form:  $P(X_1, \dots, X_n | Y) = \prod_{j=1}^n P(X_j|Y)$

How many parameters to describe  $P(X_1, \dots, X_n | Y)$ ?  $P(Y)$ ?

Without conditional independence assumption?

With conditional independence assumption

## Example :5

If the feature of Today = (Outlook is Sunny, Temp is Hot, Humidity is Normal, Windy is False), then the player will play or not?

S. No	Outlook	Temp	Humidity	Windy	Play Tennis
1	Rainy	Hot	High	False	No
2	Rainy	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	Sunny	Mild	High	False	Yes
5	Sunny	Cool	Normal	False	Yes
6	Sunny	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes
8	Rainy	Mild	High	False	No
9	Rainy	Cool	Normal	False	Yes
10	Sunny	Mild	Normal	False	Yes
11	Rainy	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Sunny	Mild	High	True	No

# Solution:

Outlook	Yes	No	
Rainy	2/9	3/5	5/14
Overcast	4/9	0/5	4/14
Sunny	3/9	2/5	5/14

Temp	Yes	No	
Hot	2/9	2/5	4/14
Mild	4/9	2/5	6/14
Cool	3/9	1/5	4/14

Humidity	Yes	No	
High	3/9	4/5	7/14
Normal	6/9	1/5	7/14

Windy	Yes	No	
True	3/9	3/5	6/14
False	6/9	2/5	8/14

$$P(Yes) = \frac{9}{14}, P(No) = 5/14$$

$$P(X_1 = Sunny) = 5/14, \quad P(X_3 = Normal) = 7/14$$

$$P(X_2 = Hot) = 4/14, \quad P(X_4 = False) = 8/14$$

Today = (Sunny, Hot, Normal, False)

Let Today = X, Sunny =  $X_1$ , Hot =  $X_2$ , Normal =  $X_3$ , False =  $X_4$

$$P(Yes/X) = \frac{P(X/Yes) P(Yes)}{P(X)} = \frac{P(X_1, X_2, X_3, X_4/Yes) P(Yes)}{P(X_1, X_2, X_3, X_4)} = \frac{P\left(\frac{X_1}{yes}\right) P\left(\frac{X_2}{yes}\right) P\left(\frac{X_3}{yes}\right) P\left(\frac{X_4}{yes}\right). P(Yes)}{P(X_1) P(X_2) P(X_3) P(X_4)}$$

$$= \frac{\frac{3}{9} \times \frac{2}{9} \times \frac{6}{9} \times \frac{6}{9} \times \frac{9}{14}}{\frac{5}{14} \times \frac{4}{14} \times \frac{7}{14} \times \frac{8}{14}} = 0.72$$

$P(No/X) = 0.157$ .  $P(Yes/X) > P(No/X)$ . Hence, Player will play Tennis.

## Example:6

---

- (a) Estimate Conditional Probability of each attributes {colour, legs, height, smelly} for the species classes: { S1, S2 } using the data given in the table.
- (b) Using these probabilities in the table estimate the probabilities values for the new instance – { Colour =Green , Legs = 2, Height = Tall and Smelly= No}

No	Colour	Legs	Height	Smelly	Species
1	White	3	Short	Yes	S1
2	Green	2	Tall	No	S1
3	Green	3	Short	Yes	S1
4	White	3	Short	Yes	S1
5	Green	2	Short	No	S2
6	White	2	Tall	No	S2
7	White	2	Tall	No	S2
8	White	2	Short	Yes	S2

## Solution:

$$P(S1) = \frac{4}{8}, P(S2) = \frac{4}{8},$$

Colour	S1	S2
White	2/4	3/4
Green	2/4	1/4

Legs	S1	S2
2	1/4	4/4
3	3/4	0/4

Height	S1	S2
Tall	3/4	2/4
Short	1/4	2/4

Smelly	S1	S2
Yes	3/4	1/4
No	1/4	3/4

## Contd..

$$P\left(\frac{S1}{New\ Instance}\right) = P(S1) \times P\left(\frac{Colour=Green}{S1}\right) \times P\left(\frac{Leg=2}{S1}\right) \times P\left(\frac{Height=Tall}{S1}\right) \times \frac{P(Smelly=no)}{S1}$$

$$= 1/2 \times 2/4 \times 1/4 \times 3/4 \times 1/4 = 0.0117$$

$$P\left(\frac{S2}{New\ Instance}\right) = P(S2) \times P\left(\frac{Colour=Green}{S2}\right) \times P\left(\frac{Leg=2}{S2}\right) \times P\left(\frac{Height=Tall}{S2}\right) \times \frac{P(Smelly=no)}{S2}$$

$$= 1/2 \times 1/4 \times 4/4 \times 2/4 \times 3/4 = 0.047$$

$$P\left(\frac{S2}{New\ Instance}\right) > P\left(\frac{S1}{New\ Instance}\right).$$

Hence, New instance belongs to species S2

# Issues with Naïve Bayes Classifier

Consider the table with Tid = 7 deleted

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				
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

## Naïve Bayes Classifier:

$$P(\text{Refund} = \text{Yes} | \text{No}) = 2/6$$

$$P(\text{Refund} = \text{No} | \text{No}) = 4/6$$

→  $P(\text{Refund} = \text{Yes} | \text{Yes}) = 0$

$$P(\text{Refund} = \text{No} | \text{Yes}) = 1$$

$$P(\text{Marital Status} = \text{Single} | \text{No}) = 2/6$$

→  $P(\text{Marital Status} = \text{Divorced} | \text{No}) = 0$

$$P(\text{Marital Status} = \text{Married} | \text{No}) = 4/6$$

$$P(\text{Marital Status} = \text{Single} | \text{Yes}) = 2/3$$

$$P(\text{Marital Status} = \text{Divorced} | \text{Yes}) = 1/3$$

$$P(\text{Marital Status} = \text{Married} | \text{Yes}) = 0/3$$

For Taxable Income:

If class = No: sample mean = 91

sample variance = 685

If class = Yes: sample mean = 90

sample variance = 25

Given  $X = (\text{Refund} = \text{Yes}, \text{Divorced}, 120\text{K})$

$$P(X | \text{No}) = 2/6 \times 0 \times 0.0083 = 0$$

$$P(X | \text{Yes}) = 0 \times 1/3 \times 1.2 \times 10^{-9} = 0$$

Naïve Bayes will not be able to  
classify X as Yes or No!

# A Simple Example:

Text	Tag
"A great game"	Sports
"The election was over"	Not sports
"Very clean match"	Sports
"A clean but forgettable game"	Sports
"It was a close election"	Not sports

Which tag does the sentence *A very close game* belong to? i.e.  $P(\text{sports} | \text{A very close game})$

Feature Engineering: Bag of words i.e use word frequencies without considering order

Using Bayes Theorem:

$$P(\text{sports} | \text{A very close game})$$

$$= P(\text{A very close game} | \text{sports}) P(\text{sports})$$

$$P(\text{A very close game})$$

We assume that every word in a sentence is **independent** of the other ones

"close" doesn't appear in sentences of sports tag, So  $P(\text{close} | \text{sports}) = 0$ , which makes product 0

## A Simple Example

Draw Naive Bayes  
Text classification

Text	Tag
"A great game"	Sports ✓
"The election was over"	Not sports ✓
'Very clean match'	Sports ✓
"A clean but forgettable game"	Sports ✓
"It was a close election"	Not sports ✓

Which tag does the sentence "A very close game" belong to? i.e.  $P(\text{sports} | \text{A very close game})$   
 Feature Engineering: Bag of words i.e. use word frequencies without considering order  
 Using Bayes Theorem:  

$$P(\text{sports} | \text{A very close game}) = \frac{P(\text{A very close game} | \text{sports}) P(\text{sports})}{P(\text{A very close game})}$$

We assume that every word in a sentence is **independent** of the other ones

$$P(\text{A very close game}) = P(A) P(\text{very}) P(\text{close}) P(\text{game})$$

$$P(\text{A very closed game} | \text{sports}) = P(\text{a} | \text{sports}) P(\text{very} | \text{sports}) \\ P(\text{close} | \text{sports}) P(\text{game} | \text{sports})$$

"close" doesn't appear in sentences of sports tag, So  $P(\text{close} | \text{sports}) = 0$ , which makes product 0

$$P(\text{sports} | \text{A very close game}) = \frac{P(\text{A very close game} | \text{sports}) P(\text{sports})}{P(\text{A very close game})}$$

$$P(\text{not sports} | \text{A very close game})$$

# Naïve Bayes for Text Classification



- Naïve Bayes is commonly used for **text classification**
- For a document with  $k$  terms  $d = (t_1, \dots, t_k)$

Fraction of documents in c

$$P(c|d) = P(c)P(d|c) = P(c) \prod_{t_i \in d} P(t_i|c)$$

- $P(t_i|c)$  = Fraction of terms from **all documents in c** that are  $t_i$

Number of times  $t_i$  appears in some document in c

$$P(t_i|c) = \frac{N_{ic} + 1}{N_c + T}$$

Laplace Smoothing

Total number of terms in all documents in c

Number of unique words (vocabulary size)

- Easy to implement and works relatively well
- **Limitation:** Hard to incorporate **additional features** (beyond words).
  - E.g., number of adjectives used.

# Laplace smoothing

---

- Laplace smoothing: we add 1 or in general constant k to every count so it's never zero.
- To balance this, we add the number of possible words to the divisor, so the division will never be greater than 1
- In our case, the possible words are ['a', 'great', 'very', 'over', 'it', 'but', 'game', 'election', 'clean', 'close', 'the', 'was', 'forgettable', 'match'].

# Apply Laplace Smoothing

Word	P(word   Sports)	P(word   Not Sports)
a	2+1 / 11+14	1+1 / 9+14
very	1+1 / 11+14	0+1 / 9+14
close	0+1 / 11+14	1+1 / 9+14
game	2+1 / 11+14	0+1 / 9+14

$$\begin{aligned}
 & P(a|Sports) \times P(very|Sports) \times P(close|Sports) \times P(game|Sports) \times \\
 & P(Sports) \\
 & = 2.76 \times 10^{-5} \\
 & = 0.0000276
 \end{aligned}$$

$$\begin{aligned}
 & P(a|Not\ Sports) \times P(very|Not\ Sports) \times P(close|Not\ Sports) \times \\
 & P(game|Not\ Sports) \times P(Not\ Sports) \\
 & = 0.572 \times 10^{-5} \\
 & = 0.00000572
 \end{aligned}$$

## Example : 7

Doc No	Text
1	I LOVED THE MOVIE
2	I HATED THE MOVIE
3	A GREAT MOVIE ,GOOD MOVIE
4	POOR ACTING
5	GREAT ACTING , A GOOD MOVIE
NEW	I HATED THE POOR ACTING

## Solution:

Doc No	Text	
1	I LOVED THE MOVIE	_POSITIVE
2	I HATED THE MOVIE	_NEGATIVE
3	A GREAT MOVIE ,GOOD MOVIE	_POSITIVE
4	POOR ACTING	_NEGATIVE
5	GREAT ACTING , A GOOD MOVIE	_POSITIVE
NEW	I HATED THE POOR ACTING	_????

Contd..

$$P(c/x)$$

$$P(+ / \text{I hated the poor acting}) =$$

$$P(- / \text{I hated the poor acting}) =$$

Based on these probabilities, we can decide the class which the new text belongs

Contd..

$P(+ | \text{I hated the acting})$

$$\text{ie } P(c_1|x) = \frac{P(x|c_1) P(c_1)}{P(x)}$$

$$= P(\text{I} | +) P(\text{hated} | +) P(\text{the} | +) P(\text{acting} | +) P(+)$$

$$P(\text{I}, +)$$

$$\xrightarrow{P(+)}$$

$$P(\text{hated}, +)$$

$$\xrightarrow{P(+)}$$

Contd..

words	positive	Negative
I	1	1
loved	1	0
the	1	1
movie	4	1
hated	0	1
a	2	0
great	2	0
Poor	0	1
acting	1	1
good	2	0

$$P(\pm | +) = \frac{1+1}{14+10}$$

$$P(\mp | -) = \frac{1+1}{6+10}$$

Contd..

"I hated the poor acting"

word	positive	negative
I	$\frac{1+1}{14+10} = 0.0833$	$\frac{1+1}{6+10} = 0.125$
hated	$\frac{0+1}{14+10} = 0.0417$	$\frac{1+1}{6+10} = 0.125$
the	$\frac{1+1}{14+10} = 0.0833$	$\frac{1+1}{6+10} = 0.125$
poor	$\frac{0+1}{14+10} = 0.0417$	$\frac{1+1}{6+10} = 0.125$
acting	$\frac{1+1}{14+10} = 0.0833$	$\frac{1+1}{6+10} = 0.125$

## Contd..

$x : I \text{ hate the Poor acting}$

$$\begin{aligned}
 P(+|x) &= (\text{ }) \times P(+) \\
 &\quad \downarrow \\
 &\quad \frac{3}{5}
 \end{aligned}$$

$$= 6.03 \times 10^{-7}$$

$$\begin{aligned}
 P(-|x) &= (\text{ }) \times P(-) : \\
 &\quad \downarrow \frac{2}{5} \\
 &= 1.22 \times 10^{-5}
 \end{aligned}$$

$\therefore$  negative class

## Example:8

Suppose we got the new message with the words '**Dear Friend**', Decide whether this new message is a normal or spam message?

i.e. Normal/ Dear, Friend = Yes or No

Note if we know  $P(\text{Normal} / \text{Dear, Friend})$  and  $P(\text{Spam} / \text{Dear, Friend})$  then we can answer the question asked

Email word	Spam	Email word	Spam
Dear	Yes	Friend	No
Friend	No	Friend	Yes
Dear	No	Dear	No
Dear	No	Lunch	No
Dear	No	Friend	No
Friend	No	Dear	No
Lunch	No	Dear	No
Friend	No	Dear	No
Lunch	No	Dear	No
Dear	Yes	Money	Yes
Money	Yes	Money	No
Money	Yes	Money	Yes

# Steps to Apply Bayes' Theorem

**Step 1- View or collect “raw” data.**

**Step 2 - Convert long data to a frequency table**

	Play		Row Total
word	normal	spam	
Dear	8	2	10
Friend	5	1	6
Lunch	3	0	3
Money	1	4	5
Column Total	17	7	24

**Step 3 - Row and column sums to get probabilities**

Email word	Spam	Email word	Spam
Dear	Yes	Friend	No
Friend	No	Friend	Yes
Dear	No	Dear	No
Dear	No	Lunch	No
Dear	No	Friend	No
Friend	No	Dear	No
Lunch	No	Dear	No
Friend	No	Dear	No
Lunch	No	Dear	No
Dear	Yes	Money	Yes
Money	Yes	Money	No
Money	Yes	Money	Yes

# Steps to Apply Bayes' Theorem



Play

	normal	spam	Row Total
word			
Dear	8	2	10
Friend	5	1	6
Lunch	3	0	3
Money	1	4	5
Column Total	17	7	24

As  $P(N/D, F) > P(S/D, F)$ , we can decide that Dear Friend is Normal message.

## Step 4 - Apply probabilities from frequency table to Bayes theorem

$$P(N/D, F) = \frac{P(D, F/N) \cdot P(N)}{P(D, F)} = \frac{P(D/N) \cdot P(F/N) \cdot P(N)}{P(D) \cdot P(F)} = \frac{\left(\frac{8}{17}\right) \cdot \left(\frac{5}{17}\right) \cdot \left(\frac{17}{24}\right)}{\left(\frac{10}{24}\right) \cdot \left(\frac{6}{24}\right)} = \frac{0.098}{0.104} = 0.9423$$

$$P(S/D, F) = \frac{P(D, F/S) \cdot P(S)}{P(D, F)} = \frac{P(D/S) \cdot P(F/S) \cdot P(S)}{P(D) \cdot P(F)} = \frac{\left(\frac{2}{7}\right) \cdot \left(\frac{1}{7}\right) \cdot \left(\frac{7}{24}\right)}{\left(\frac{10}{24}\right) \cdot \left(\frac{6}{24}\right)} = \frac{0.012}{0.104} = 0.1153$$

## Example continued:

Suppose we got the new message contains the word '**Lunch Money Money Money Money**' , Decide whether this new message is a normal or spam message?

Email word	Spam	Email word	Spam
Dear	Yes	Friend	No
Friend	No	Friend	Yes
Dear	No	Dear	No
Dear	No	Lunch	No
Dear	No	Friend	No
Friend	No	Dear	No
Lunch	No	Dear	No
Friend	No	Dear	No
Lunch	No	Dear	No
Dear	Yes	Money	Yes
Money	Yes	Money	No
Money	Yes	Money	Yes

# Steps to Apply Bayes' Theorem



Play

	normal	spam	Row Total
word			
Dear	8	2	10
Friend	5	1	6
Lunch	3	0	3
Money	1	4	5
Column Total	17	7	24

We can observe that we have to classify any message with Lunch as Normal message, no matter how many times we see the word Money and that's the problem.

To work around this problem add 1 count to the frequency table to each word(Laplace smoothing)

## Step 4 - Apply probabilities from frequency table

$$P(N) \cdot P(L/N) \cdot P(M/N)^4 = \left(\frac{17}{24}\right) \cdot \left(\frac{3}{17}\right) \cdot \left(\frac{1}{17}\right)^4 = 0.0000015$$

$$P(S) \cdot P(L/S) \cdot P(M/S)^4 = \left(\frac{7}{24}\right) \cdot \left(\frac{0}{7}\right) \cdot \left(\frac{4}{7}\right)^4 = 0$$

# Steps to Apply Bayes' Theorem



Play

	normal	spam	Row Total
word			
Dear	8+1	2+1	12
Friend	5+1	1+1	8
Lunch	3+1	0+1	5
Money	1+1	4+1	7
Column Total	21	11	32

As  $P(S/L, M^4) > P(N/L, M^4)$ , we can decide that  
**Lunch Money Money Money Money is Spam**  
message.

## Step 4 - Apply probabilities from frequency table

$$P(N) \cdot P(L/N) \cdot P(M/N)^4 = \left(\frac{21}{32}\right) \cdot \left(\frac{4}{21}\right) \cdot \left(\frac{2}{21}\right)^4 = 0.00001$$

$$P(S) \cdot P(L/S) \cdot P(M/S)^4 = \left(\frac{11}{32}\right) \cdot \left(\frac{1}{11}\right) \cdot \left(\frac{5}{11}\right)^4 = 0.00133$$

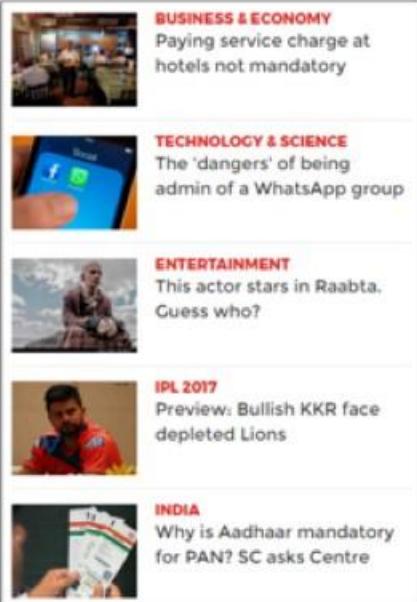
# Naïve Bayes Classifier Applications

innovate

achieve

lead

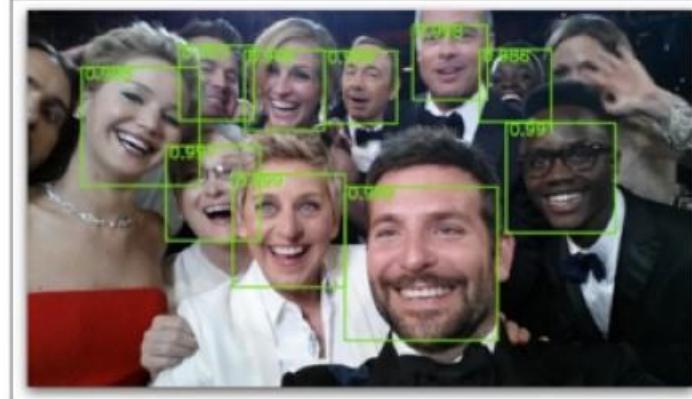
## Categorizing News



## Email Spam Detection



## Face Recognition



## Sentiment Analysis



# Naive Bayes Classifier

---

- ✓ Along with decision trees, neural networks, one of the most practical learning methods.
- ✓ When to use
  - ✓ Moderate or large training set available
  - ✓ Attributes that describe instances are conditionally independent given classification
- ✓ Successful applications:
  - ✓ Diagnosis
  - ✓ Classifying text documents

- Require initial knowledge of many probabilities
  - Often estimated based on background knowledge, previously available data, and assumptions about the form of the underlying distributions.
- Significant computational cost required to determine the Bayes optimal hypothesis in the general case (linear in the number of candidate hypotheses)

# Practice problems

## Problem 1

---

Two sets of candidates are competing for the positions on the board of directors of a company. The probabilities that the first and the second sets will win are 0.6 and 0.4 respectively. If the first set wins, the probability of introducing a new product is 0.8, and the corresponding probability if the second set wins is 0.3. What is the probability that the product will be introduced?

# Problem 1

## Solution:



Let  $E_1, E_2$  denote the events that the First and the second sets of candidates win and  $E$  be the Event of introducing a new product.

$$P(E_1) = 0.6; P(E_2) = 0.4; P(E/E_1) = 0.8; P(E/E_2) = 0.3;$$

By Addition theorem of probability

$$\begin{aligned} P(E) &= P(E_1 \cap E) + P(E_2 \cap E) = P(E_1)P\left(\frac{E}{E_1}\right) + P(E_2)P\left(\frac{E}{E_2}\right) = 0.6 \times 0.8 + 0.4 \times 0.3 \\ &= 0.48 + 0.12 = 0.6. \end{aligned}$$

## Problem 2

---

In a bolt factory, Machines A, B, C manufacture respectively 25%, 35% and 40 % of the total. Of their output 5, 4, 2 percent are known to be defective bolts.

A bolt is drawn at random from the product and is found to be defective. What are the probabilities that it was manufactured by

- (i) Machine A
- (ii) Machine B or C

## Problem 2

### Solution:

Let  $E_1, E_2$  and  $E_3$  denote respectively the events that the bolt selected at random is manufactured by the machines A, B and C respectively.

Let E denote the event that it is defective.

(i) The probability that the defective bolt is manufactured by Machine A is:

$$P(E_1) = 0.25, P(E_2) = 0.35, P(E_3) = 0.40$$

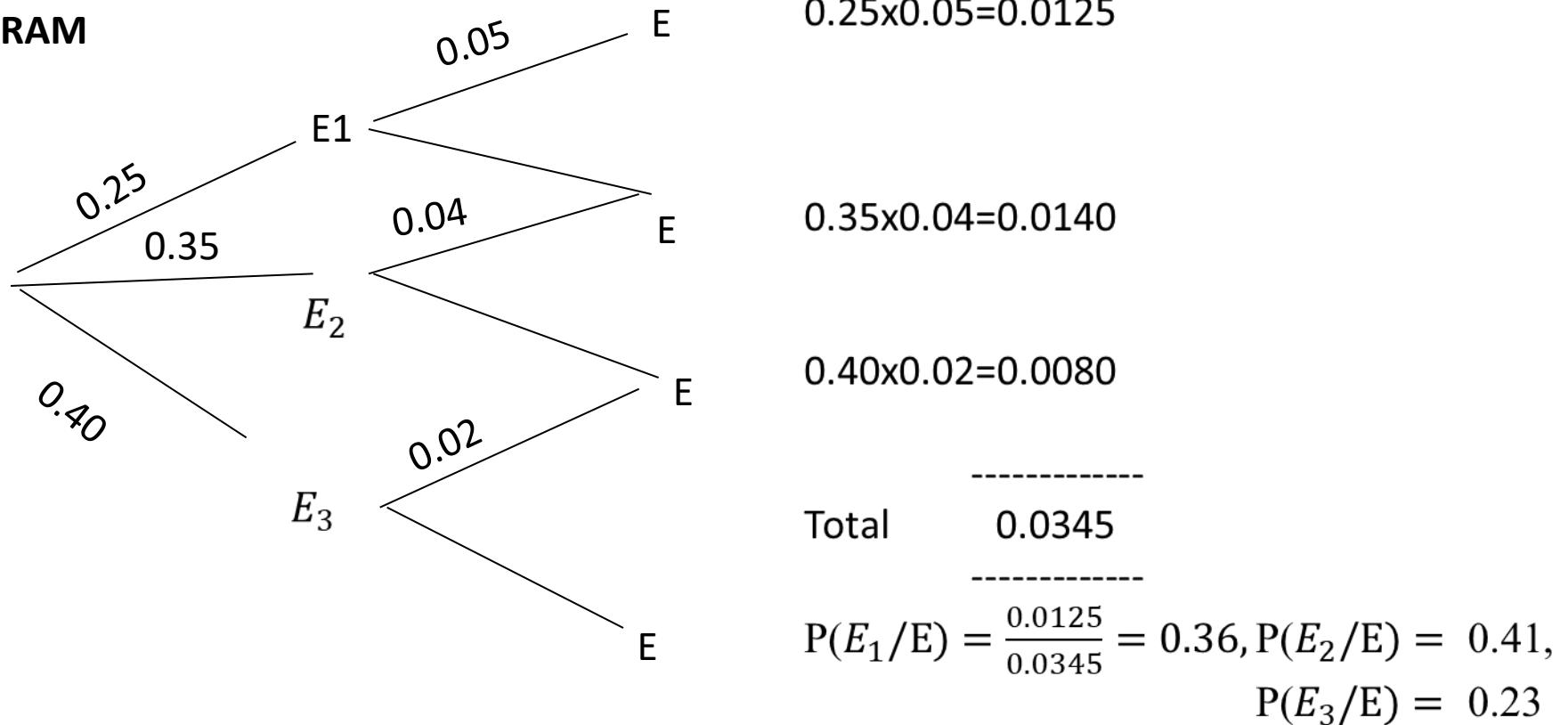
$$P(E/E_1) = 0.05, P(E/E_2) = 0.04, P(E/E_3) = 0.02,$$

$$P(E) = 0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02 = 0.0345.$$

$$P(E_1/E) = \frac{P(E_1)P(\frac{E}{E_1})}{\sum P(E_1)P(\frac{E}{E_1})} = \frac{0.0125}{0.0345} = 0.36$$

(ii) Similarly,  $P(E_2/E) = 0.41, P(E_3/E) = 0.23$ , Hence Defective bolt chosen at random is Manufactured by B or C =  $P(E_2/E) + P(E_3/E) = 0.41 + 0.23 = 0.64$ .

## TREE DIAGRAM



## Problem 3

---

The results of an investigation by an expert on a fire accident in a skyscraper are summarized below:

- (i)  $P(\text{There should have been short circuit})=0.8$
- (ii)  $P(\text{LPG cylinder explosion})$
- (iii) Chance of fire accident is 30% given a short circuit and 95% given an LPG explosion.

Based on these, What do you think is the most probable causes of fire? Statistically justify your answer.

## Solution:

Let  $E_1$  : Short circuit ;  $E_2$ : LPG explosion ; E: Fire accident

Then,

$$P(E_1) = 0.8; P(E_2) = 0.2; P(E/E_1) = 0.30; P(E/E_2) = 0.95$$

By Baye's Rule:

$$P(E_1/E) = \frac{P(E_1) \cdot P(E/E_1)}{P(E_1) \cdot P(E/E_1) + P(E_2) \cdot P(E/E_2)} = \frac{0.8 \times 0.30}{0.8 \times 0.30 + 0.2 \times 0.95} = \frac{0.240}{0.240 + 0.190} = \frac{24}{43}$$

$$P(E_2/E) = \frac{P(E_2) \cdot P(E/E_2)}{P(E_1) \cdot P(E/E_1) + P(E_2) \cdot P(E/E_2)} = \frac{0.190}{0.430} = \frac{19}{43}$$

(OR)

$$P(E_2/E) = 1 - P(E_1/E) = 1 - \frac{24}{43} = \frac{19}{43}$$

## Problem 4

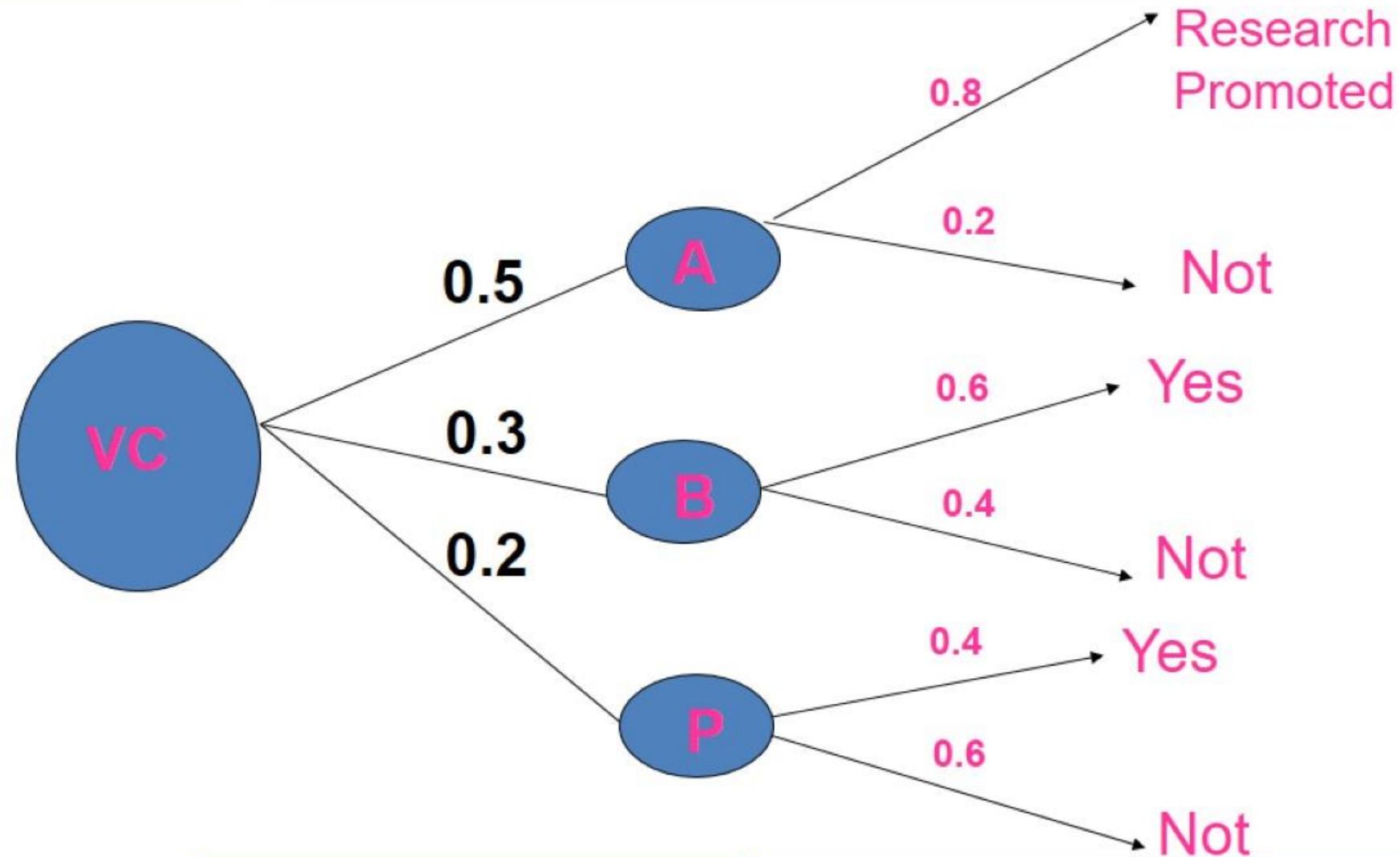
The chances that an academician, a business man and a politician becoming Vice Chancellor of an university are 0.5, 0.3 and 0.2 respectively. The probability that research work will be promoted in the university by these 3 gentlemen are respectively are 0.8, 0.6 and 0.4. It is found Research work has been promoted by the university. What is the chance that an academician has become the VC?

**Solution:** Let X: “Research work is promoted”

$$\text{Clearly, } P(X) = 0.5 \times 0.8 + 0.3 \times 0.6 + 0.2 \times 0.4 = 0.66$$

$$\text{Now to find } P[\text{“An Academician is VC”} / \text{“Research work is promoted i.e. event X”}] = \frac{0.5 \times 0.8}{0.66} = 0.6061$$

# Tree Diagram



## Problem 5

---

Four technicians regularly make repairs when breakdowns occur on an automated production line.

- Janet, who services 20% of the breakdowns, makes an incomplete repair 1 time in 20;
- Tom, who services 60% of the breakdowns, makes an incomplete repair 1 time in 10;
- Georgia, who services 15% of the breakdowns, makes an incomplete repair 1 time in 10;
- Peter, who services 5% of the breakdowns, makes an incomplete repair 1 time in 20.

For the next problem with the production line diagnosed as being due to an initial repair that was incomplete, what is the probability that this initial repair was made by Janet?

## Solution:

---

Let  $A$  be the event that the initial repair was incomplete,  $B_1$  that the initial repair was made by Janet,  $B_2$  that it was made by Tom,  $B_3$  that it was made by Georgia, and  $B_4$  that it was made by Peter.

$$P(B_1|IA) = \frac{(0.20)(0.05)}{(0.20)(0.05) + (0.60)(0.10) + (0.15)(0.10) + (0.05)(0.05)}$$
$$= \mathbf{0.114}$$

and it is of interest to note that although Janet makes an incomplete repair only 1 out of 20 times, namely, 5% of the breakdowns she services, more than 11% of the incomplete repairs are her responsibility.

## Problem 6

---

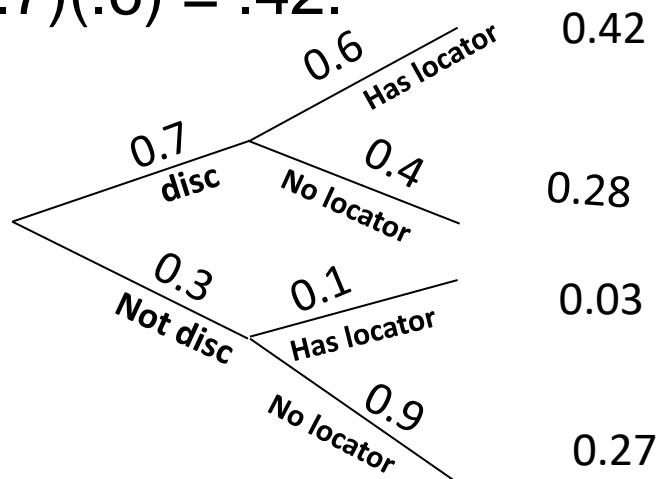
Seventy percent of the light aircraft that disappear while in flight in a certain country are subsequently discovered. Of the aircraft that are discovered, 60% have an emergency locator, whereas 90% of the aircraft not discovered do not have such a locator. Suppose that a light aircraft has disappeared.

- (a) If it has an emergency locator, find the probability that it will not be discovered.
- (b) If it doesn't have a locator, find the probability that it will be discovered.

## Solution:

The tree diagram below shows the probability for the four disjoint options;  
 $P(\text{the flight is discovered and has a locator})$

$$\begin{aligned}
 &= P(\text{discovered})P(\text{locator} \mid \text{discovered}) \\
 &= (.7)(.6) = .42.
 \end{aligned}$$



$$\begin{aligned}
 \text{a. } P(\text{not discovered} \mid \text{has locator}) &= \frac{P(\text{not discovered} \cap \text{has locator})}{P(\text{has locator})} = \frac{0.03}{0.03+0.42} = 0.067 \\
 \text{b. } P(\text{discovered} \mid \text{no locator}) &= \frac{P(\text{discovered} \cap \text{no locator})}{P(\text{no locator})} = \frac{0.28}{0.28+0.27} = 0.509
 \end{aligned}$$

## Problem 7

---

Two firms V and W consider bidding on a road-building job, which may or may not be awarded depending on the amounts of the bids. Firm V submits a bid and the probability is  $\frac{3}{4}$  that it will get the job provided firm W does not bid. The probability is  $\frac{3}{4}$  that W will bid, and if it does, the probability that V will get the job is only  $\frac{1}{3}$ . (a) what is the probability that V will get the job? (b) If V gets the job, what is the probability that W did not bid?

## Solution:

---

Given:  $P(V/W') = 3/4$ ,  $P(W) = 1/3$ ,  $P(V/W) = 1/3$ ,  $P(W') = 2/3$

(a)  $V = (V \cap W) \cup (V \cap W') = P(V) = P(V \cap W) + P(V \cap W')$

$$P(V) = P(V/W) P(W) + P(V/W') P(W') = 11/18$$

(b)  $P(W'/V) = P(V/W') P(W') / P(V) = 9/11$

# Glossary

- Conditional probability
- Total Probability
- Bayes theorem
- MAP Hypothesis
- ML Estimate
- Naïve Bayes theorem
- NB with Laplace correction

# IMP Note to Self

---



*Thank you*