BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction to Statistical Methods

ISM Team

Session 4
(24th / 25th May 2025)

Baye’s theorem and Naïve Bayes theorem

Conditional Probability

$P(A) = ?$ → Prior

$P(A|B)$ → Post

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

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

Independence

$P(A \cap B) = P(A) \times P(B)$

$P(A) = P(A|B)$

$P(A|E_1) \quad P(E_1)$

✓ Multiplication Principle

$P(A \cap B) = P(A|B) \cdot P(B)$

$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \ldots + P(A_n \cap B)$

$= P(A_1|B) \cdot P(B) + \ldots$

Contact Session 4

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

Agenda

- Baye’s theorem
- Problems on Baye’s theorem
- Naïve Bayes theorem

BAYE'S THEOREM:

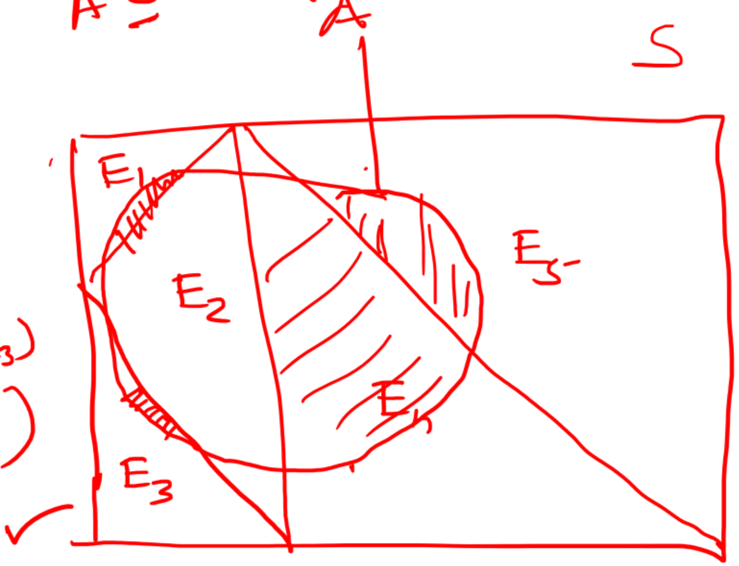
Suppose that E_1, E_2 , ---------E_n are mutually exclusive events of a sample space " S "

such that P (Ei) > 0 for i = 1,2,3, ---------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

for i = 1,2,3………n is given by $P(E_i/A) = \dfrac{P(E_i)P(A/E_i)}{\sum_{i}^{n} P(E_i)P(A/E_i)}$

$A \subseteq \cup E_i$



$$\underline{\underline{P(E_1|A)}} = \frac{P(E_1)\cdot P(A|E_1)}{\left(P(E_1)\cdot P(A|E_1) + P(E_2)\cdot P(A|E_2) + P(E_3)\cdot P(A|E_3) + P(E_4)\cdot P(A|E_4) + P(E_5)\cdot P(A|E_5)\right)}$$

$$= \frac{P(E_1)\cdot P(A|E_1)}{P(A)} = \frac{P(E\cap A)}{P(A)}$$

Proof:

$P(E_i|A) = \boxed{\dfrac{P(E_i)\cdot P(A|E_i)}{\sum P(E_i)\cdot P(A|E_i)}}$ $\quad \dfrac{P(A\cap E_i)}{P(A)}$

$P(E_i|A) = \dfrac{P(E_i \cap A)}{P(A)}$

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

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

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

$A = (A \cap E_1) \cup (A \cap E_2) \ldots\ldots\ldots\ldots \cup (A \cap E_n)$ (Distributive law)

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

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

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

$(A\cap E_i) \cap (A\cap E_j) = \varphi$ for $i \neq j$

Proof:(continuation)

- From the definition of conditional probability

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

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

$$P(E_i / A) = \frac{P(E_i)P(A/E_i)}{\sum_{i}^{n} P(E_i)P(A/E_i)}$$

Example:

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:

E_1 = X will be Manager E_2 = Y will be manger E_3 = Z will be Manger

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.

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

$P(E_3|A)=\frac{12}{23}$

(i) Using Addition theorem of Probability

$$P(A) = P(E_1 \cap A) + P(E_2 \cap A) + P(E_3 \cap A)$$

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

(ii) Using Baye's Theorem,

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

$$\frac{\frac{4}{9}\times\frac{3}{10}}{\frac{23}{45}} = \frac{12}{90}\times\frac{45}{23} = \frac{6}{23}$$

$$P(E_2|A) = \frac{P(E_2)\cdot P(A|E_2)}{P(A)} = \frac{\frac{2}{9}\times\frac{1}{2}}{23/45} = \frac{2}{18}\times\frac{45}{23}$$

Example:

- 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

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

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

$$P(R) = P(R \cap F) + P(R \cap M)$$

$$= P(R|F) \cdot P(F) + P(R|M) \cdot P(M)$$

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

$$=$$

Example

- ➢ 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%.

- o Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

$$P(D^c|S) = \frac{P(S|D^c)\cdot P(D^c)}{P(S)}$$

$P(S) = 0.5$

$P(D|S) = 0.99$

$P(D^c|S) = 0.01$

$P(S^c) = 0.5$

$P(D|S^c) = 0.05$

$P(D^c|S^c) = 0.95$

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, given.

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

Hence by the Bayes's formula

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

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

$P(A \mid B) = 0.99$

$$P(A^c \mid B) = \frac{P(A^c \cap B)}{P(B)}$$

$$= \frac{P(B \mid \cancel{A}^c) \cdot P(A^c)}{P(B)}$$

Example:

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

- Note: If we know P(Yes/Sunny) and P(No/Sunny) then we can answer the question asked

	Weather	Play
1	Sunny	No
2	Overcast	Yes
3	Rainy	Yes
4	Sunny	Yes
5	Sunny	Yes
6	Overcast	Yes
7	Rainy	No
8	Rainy	No
9	Sunny	Yes
10	Rainy	Yes
11	Sunny	No
12	Overcast	Yes
13	Overcast	Yes
14	Rainy	No

Solution

- Convert long data into frequency table

$P(Y \mid sunny)$

$P(N \mid sunny)$

Weather	No	Yes	
Sunny	2	3	= 5
Overcast	0	4	= 4
Rainy	3	2	= 5
	5	9	= 14

- 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

continuation

Weather	No	Yes	Row Total	
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		

- Apply Probabilities from frequency table to Bayes Theorem

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

- P(Yes | Sunny) > P(No | Sunny) . Hence, The Player will play the game.

Baye's Theorem

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

Hypothesis → Play of tennis

p(h = yes)

p(h = No)

$$p(D|h) \times p(h)$$

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

p(h = 1)

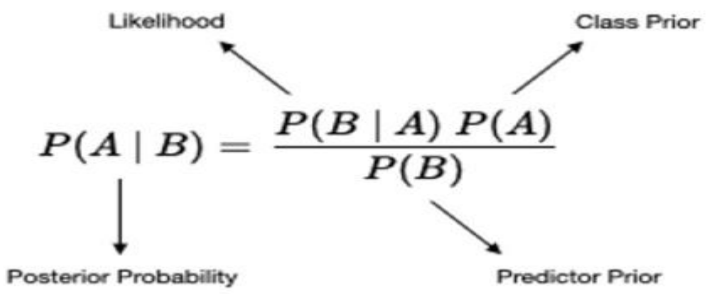
p(h = 2)

p(h = 3)

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=ck,

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



Choosing hypothesis

Generally want the most probable hypothesis given the training data

Maximum a posteriori hypothesis h_{MAP}:

Choose

Max. Post Prob

P(Y|D)

P(N|D)

$$h_{MAP} = \arg\max_{h \epsilon H} P(h|D)$$

$$= \arg\max_{h \epsilon H} \frac{P(D|h).P(h)}{P(D)}$$

$$= \arg\max_{h \epsilon H} P(D|h).P(h)$$

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

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

$P(h = h_1) = P(h = h_2) = P(h = h_3)$ - - - - Assumption

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.

X Y

X_1	X_2	X_3	X_4	X_5	
wt.	Age	blood Pressure	Sugar	stress	hypothesis = Yes = No.

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*

$P(A \cap B) = P(A) \times P(B)$

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

$$\boxed{P(X \mid Y, Z) = P(X \mid Z)}$$

$$P(X \mid Y \cap Z) = P(X \mid Z)$$

Example:

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

$$P(X_1, X_2, X_3, X_4 \mid Y) = P(X_1 \cap X_2 \cap X_3 \cap X_4 \mid Y)$$

$$= P(X_1 \mid Y) \cdot P(X_2 \mid Y) \cdot P(X_3 \mid Y) \cdot P(X_4 \mid Y)$$

Applying Conditional Independence

Naive Bayes assumes X_i are conditionally independent given Y.

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

$$P(X_1, X_2 | Y) = P(X_1 | X_2, Y)\, P(X_2 | Y)$$
$$= P(X_1 | Y)\, P(X_2 | Y)$$

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

How many parameters to describe $P(X_1, \ldots, X_n \mid Y)$? $P(Y)$?

Without conditional independence assumption?

With conditional independence assumption

Example

- If the feature of Today $X =$ (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 Y
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

$P(Yes \mid X)$

$P(Yes \mid x_1 = Sunny \cap x_2 = hot \cap x_3 = Normal \cap x_4 = False)$

$P(No \mid X)$

Solution:

X_1

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

X_2

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

X_3

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

X_4

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

continuation

$P(Y|X)$

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

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

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

X = Today = (Sunny, Hot, Normal, False)

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

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

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

$$P(No|X) = \frac{P(X|No).P(No)}{P(X)}$$

P(No/X) = 0.~~1~~57. 0.28 P(Yes/X) > P(No/X) . Hence, Player will play Tennis.

Example:

(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}

Example:

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

- Estimate Conditional Probability of each attributes {colour, legs, height, smelly} for the species classes: { S1, S2 } using the data given in the table.
- Using these probabilities in the table estimate the probabilities values for the new instance –

  {Colour =Green , Legs = 2, Height = Tall and Smelly= No}

Solution:

$P(S1) = 4/8$, $P(S2) = 4/8$

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

Height	S1	S2
Tall	¾	2/4
Short	¼	2/4

Legs	S1	S2
2	¼	4/4
3	¾	0/4

Smelly	S1	S2
Yes	¾	¼
No	¼	3/4

new instance(NI) – { Colour =Green , Legs = 2, Height = Tall and Smelly= No}

P(S1/NI)=P(S1).P(S1/Colour=Green).P(S1/Legs=2).P(S1/Height=Tall)P(S1/Smelly= No).

= (1/2).(2/4).(1/4).(3/4).(1/4) = 0.0117

P(S2/NI)=P(S2).P(S2/Colour=Green).P(S2/Legs=2).P(S2/Height=Tall).P(S2/Smelly= No).

= (1/2).(1/4).(4/4).(2/4).(3/4) = 0.047

P(S2/NI) > P(S1/NI)

Hence New Instance belongs to 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	Yes	Divorced	220K	**No**
9	No	Married	75K	**No**
10	No	Single	90K	**Yes**

Naïve Bayes Classifier:

$P(x_1 | y) \cdot P(x_2 \cdot y) \cdot P(x_3 | y) . P(x_4 | y) . P(y$

1.P(Refund = Yes | No) = 2/6

2.P(Refund = No | No) = 4/6

3.P(Refund = Yes | Yes) = 0

4.P(Refund = No | Yes) = 1

5.P(Marital Status = Single | No) = 2/6

6.P(Marital Status = Divorced | No) = 0

7.P(Marital Status = Married | No) = 4/6

8.P(Marital Status = Single | Yes) = 2/3

9.P(Marital Status = Divorced | Yes) = 1/3

10.P(Marital Status = Married | Yes) = 0/3

Given X = (Refund = Yes, Divorced, 120K)

$P(X \mid No) = 2/6 \times 0 \times 0.0083 = 0$

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

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(sports| *A very close game*)

Using Bayes Theorem:

$$P(\text{sports} \mid \textit{A very close game}) = \frac{P(\textit{A very close game} \mid \text{sports})\, P(\text{sports})}{P(\textit{A very close game})}$$

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

$P(\text{close} \mid \text{sport}) = 0$

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

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

$P(\text{A very close game} \mid \text{sports}) = P(\text{a} \mid \text{sports})\, P(\text{very} \mid \text{sports})$

$\underline{P(\text{close} \mid \text{sports})}\, P(\text{game} \mid \text{sports})$

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

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'].

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

Laplace Smoothing

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

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.

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

Num

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

Num

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

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 _????

$P(c/x)$

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

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

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

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

$$\text{ie} \quad P(c_1 \mid x) = \frac{P(x \mid c_1)\, P(c_1)}{\cancel{P(x)}}$$

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

$$\frac{P(\text{I}, +)}{P(+)} \qquad \frac{P(\text{hated}, +)}{P(+)}$$

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

$$P(I \mid +) = \frac{1+1}{14+10}$$

$$P(I \mid -) = \frac{1+1}{6+10}$$

"I hated the poor acting"

	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	$\frac{1+1}{14+10} = 0.0833$	$\frac{1+1}{6+10} = 0.125$

$P(+/x)$

$= (\quad)(\quad)(\quad)(\quad)(\quad) \times P(+)$

$\downarrow 3/5$

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

$P(-/x)$

$= (\quad)(\quad)(\quad)(\quad)(\quad) P(-1)$:

$\downarrow 2/8$

$= 1.22 \times 10^{-5}$

$\therefore$ Negative class

Example

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 (Normal/ Dear, Friend) and P(Spam/ 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		
	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

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 word	normal	spam	Row Total
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

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

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

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

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

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

Steps to Apply Bayes Theorem

Play word	normal	spam	Row Total
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

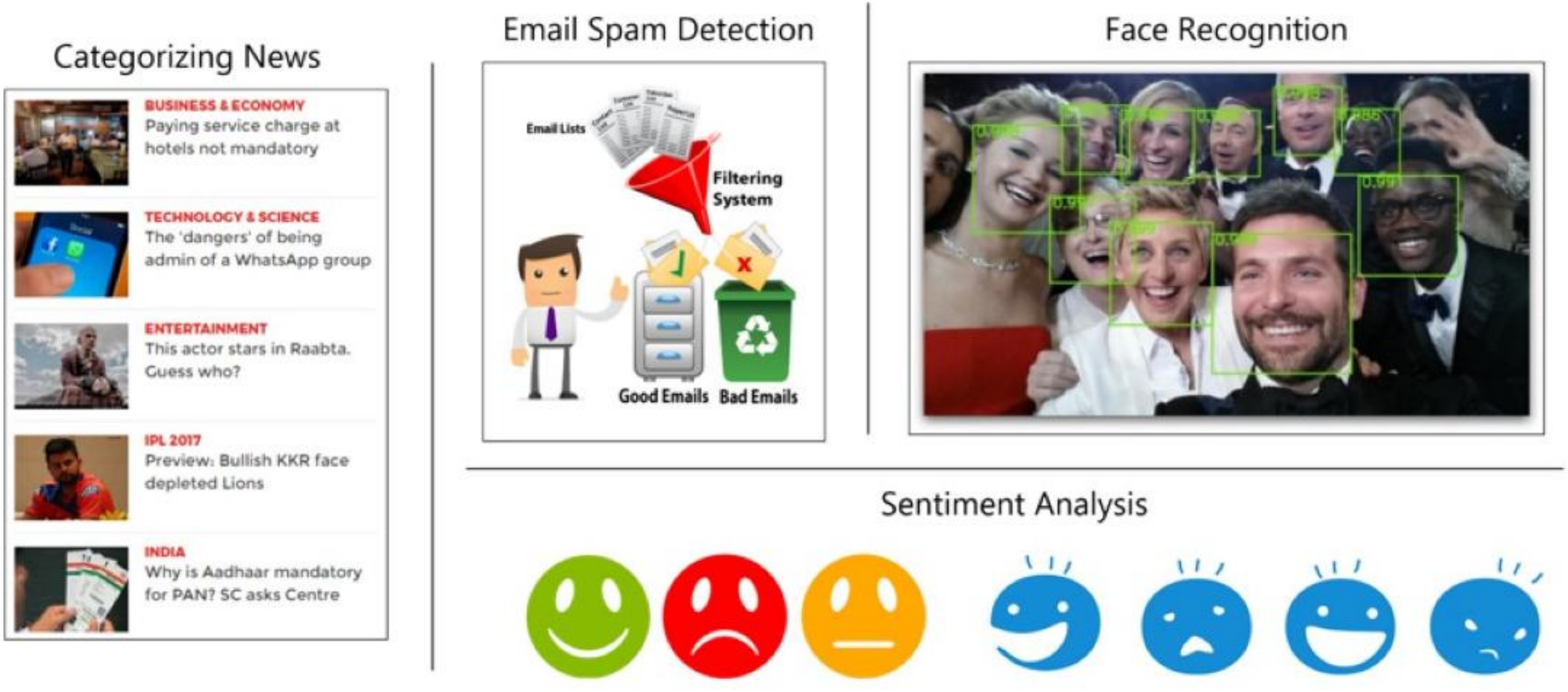
Step 4 - Apply probabilities from frequency table

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

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

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

Naïve Bayes Classifier Applications



Practical Issues of Bayesian learning

- ✓ 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

Practical Issues of Bayesian learning

- 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)

Glossary

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

Practice Problems 1

Two set 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?

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

P(E) = P$(E_1 \cap E) + \mathrm{P}(E_2 \cap \mathrm{E}) = \mathrm{P}(E_1)P(\frac{E}{E_1}) + \mathrm{P}(E_2)P(\frac{E}{E_2}) =$ 0.6x0.8+0.4x0.3

= 0.48+0.12=0.6.

Practice Problems 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

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,

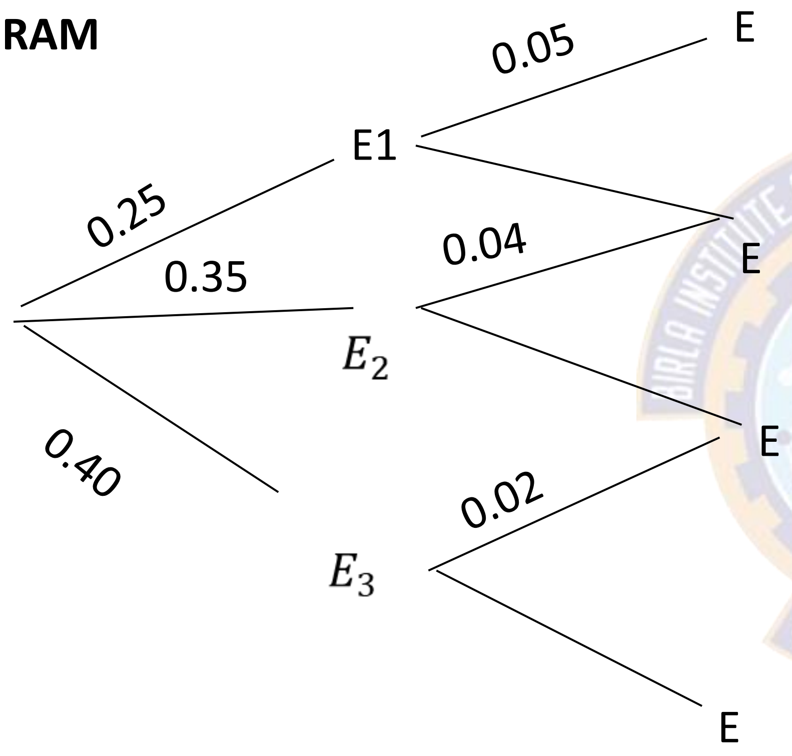
$P(E)$= 0.25 x 0.05 + 0.35x0.04 + 0.40 x 0.02 = 0.0345.

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

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

Conti..

TREE DIAGRAM



0.25x0.05=0.0125

0.35x0.04=0.0140

0.40x0.02=0.0080

Total 0.0345

$$P(E_1/E) = \frac{0.0125}{0.0345} = 0.36, P(E_2/E) = 0.41,$$

$$P(E_3/E) = 0.23$$

Practice Problems 3

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

(i) P(There should have been short circuit)=0.8

(ii) P(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:

$$P(E_1/E) = \frac{P(E_1)\,.P(E/E_1)}{P(E_1).P(E/E_1)+P(E_2)P(E/E_2)} = \frac{0.8 \times 0.30}{0.8\times0.30 + 0.2\times0.95} = \frac{0.240}{0.240+0.190} = \frac{24}{43}$$

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

(OR)

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

Practice Problems 4

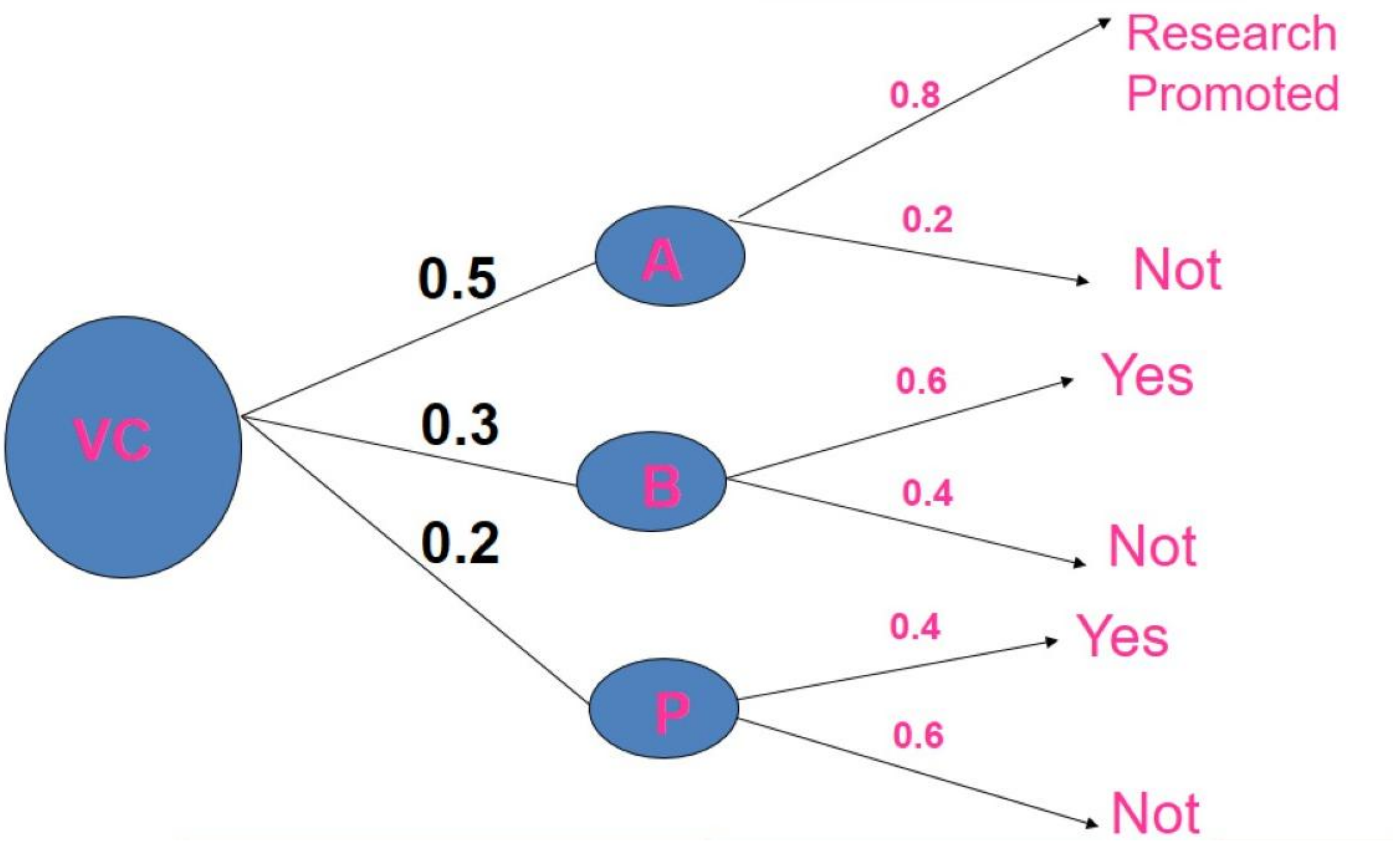
- The chances that an academician, a business man and a politician becoming Vice Chancellor of an university are 0.5, 0.3 and 0.2 respectively. The probability that research work will be promoted in the university by these 3 gentlemen are respectively are 0.8, 0.6 and 0.4. It is found Research work has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

Let X: " Research work is promoted"

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

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

Conti..



VC
0.5
0.3
0.2
A
B
P
0.8
0.2
0.6
0.4
0.4
0.6
Research Promoted
Not
Yes
Not
Yes
Not

Practice Problems 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.

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

$$= \mathbf{0.114}$$

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.

Thank You