Naive Bayes Method (Bayes Theorem).

Registration No. .... Question ....

Write on both sides of the paper

Do not write in this margin

- Classification using Bayes Theorem is different from the decision tree approach.
- It based on a hypothesis that the given data belongs to a particular class.
- In this Theorem probability is calculated for the hypothesis to be true.
- Bayes theorem is based on probabilities and so it is important to define some notation in the beginning.

P(A) refers to the probability of occurrence of event A. while P(A/B) refers to the conditional probability of event A given that event B has already occurred.

Bayes theorem is defined as follows:

$$P(A/B) = P(B/A)\,P(A) / P(B). \quad — (1)$$

let us ~~first~~ prove this theorem.
we already know that

$$P(A/B) = P(A \& B) / P(B) \quad — (2)$$

(It is the probability that the next event will be A and B has already happened)

Similarly =

$$P(B/A) = P(B \& A) / P(A) \quad — 3.$$

From equation (3):

Thus,

$$P(B \& A) = P(B/A) * P(A)$$

By putting this value of $P(B \& A)$ in equation (2), we get

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

- Predicting Class of X out of three given classes C_1, C_2 and C_3.

- Hypothesis is that event X has already occurred.
- Thus, we learn to calculate $P(C_i/X)$, conditional probability of class being C_i when X has already occurred.

According to Bayes theorem:

$$P(C_i/X) = \frac{P(X/C_i)\,PC_i}{P(X)}$$

In this equation:

$P(C_i|X)$ is the conditional probability of class being C_i when X has already occurred or it is probability that X belongs to class C_i

$P(X|C_i)$ - Is the conditional probability of record being X when it is known that output class is C_i

$P(C_i)$ is probability that object belongs to class C_i

Registration No. .... Question ....

Write on both sides of the paper

Do not write in this margin

P(X) is the probability of occurrence of record X.

Here, we have already made the hypothesis that X has already occurred so P(X) is 1. So we have to calculate P(X|C₁) and P(C₁) in order to find the required value.

Example:
Then we have to calculate the probability for C_i when X has already occurred. Thus it is

$$P(C_1 \mid \text{Yes, No, Male, Yes, A}) = P(\text{Yes, No, Male, Yes, A} \mid C_1) * P(C_i).$$

Let us first calculate the probability of each class, i.e. $P(C_i)$.
- From the table, there are three classes i.e. I, II and III
- There are total 10 instances in the given dataset

Class I = 3 instances
Class II = 3 instances
Class III = 4 instances

} Total

Probability of Class I, i.e... $P(I) = 3/10 = 0.3$.

II $P(II) = 3/10 = 0.3$.

III $P(III) = 4/10 = 0.4$.

Registration No. ....

in this margin

Write on both sides of the paper

Now let us calculate $P(X|C_1)$ i.e

$P(\text{Yes, No, Male, Yes, A} | C_1)$.

$P(\text{Yes, No, Male, Yes, A} | C_i) = P(\text{Owns Home} = \text{Yes} | C_i) *$

$P(\text{Married} = \text{No} | C_i) * P(\text{Gender} = \text{Male} / C_1) * P(\text{Employed} =$

$\text{Yes} / C_i) * P(\text{Credit rating} = A / C_1)$.

Thus we need to calculate.

$P(\text{Owns Home} = \text{Yes} | C_i)$,

$P(\text{Married} = \text{No} / C_i)$

$P(\text{Gender} = \text{Male} | C_i)$

$P(\text{Employed} = \text{Yes} | C_1)$

$P(\text{Credit rating} = A | C_i)$

Own Home	Married	Gender	Employed	Credit Rating	Class
No	No	Female	Yes	A	1
No	No	Female	Yes	B	1
Yes	No	Female	Yes	A	1
Yes	No	Male	Yes	A	? (To fi...
1/3		0	1	2/3.	

Registration No. .... Question ....

Do not write in this margin

Write on both sides of the paper

Probability of having (Yes, NO, Male, Yes, A) Attribute given the risk class I.

Own Home	Married	Gender	Employed	Credit rating	Class
Yes	Yes	Male	Yes	A	II
Yes	NO	male	NO	B	III
NO	NO	Male	NO	B	II
Yes	NO	Male	Yes	A	I
2/3	2/3	1	1/3	1/3	Probability (Yes, no, male, yes, A) Attribute value given risk value II
Yes	Yes	Female	Yes	B	III
NO	Yes	Female	Yes	B	III
NO	Yes	Female	Yes	A	III
Yes	Yes	Female	Yes	A	III
Yes	NO	male	Yes	A	To be found
2/4 = 1/2	0	0	1	2/4 = 1/2	Attribute value given the risk class III

Thus, $P(X|I) = \frac{1}{3} \times 1 \times 0 \times 1 \times \frac{2}{3} = 0$.

$P(X|II) . \frac{2}{3} \times \frac{2}{3} \times 1 \times \frac{1}{3} \times \frac{1}{3} = \frac{4}{81}$

$P(x|III) . \frac{1}{2} \times 0 \times 0 \times 1 \times \frac{1}{2} = 0$.

$P(I / Yes, No, Male, Yes, A) = P(Yes, No, Male, Yes, A / I) \times P(I)$

$0 \times 0.3 = 0$

Registration No. .... Question ....

Write on both sides of the paper

Do not write in this margin

$P(II / Yes, No, Male, Yes, A) = P(Yes, No, Male, Yes, A / II) * P(II)$

$= 4/81 \times 0.3 = 0.0148$

$P(III / Yes, No, Male, Yes, A) = P(Yes, No, Male, Yes, A / III)$

$P(III) = 0 * 0.4 = 0.$

Therefore X is assigned to class III. It is very unlikely in real life datasets that the probability of class comes out to be 0 as in this exam.

Outlook	Temperature	Humidity	Windy	Play
Sunny = 5	hot = 4	high = 7	true = 6	Yes = 9
Overcast = 4	mild = 6	normal = 7	false = 8	no = 5
rainy = 5	Cool = 4			

From the dataset, we can observe that when the Outlook is Sunny, Play is not held on 3 out of 5 days.
- Therefore we can conclude that there are 60% chances that play will not be held when it is a sunny day.

- Naive Bayes Theorem is based on Probability, so we need to calculate probabilities for each occurrence in the shown

Registration No. .... Question ....

Do not write in this margin

Write on both sides of the paper

	Play			Play			Play			Play	
Outlook	Yes	NO	Temp	Yes	NO	Humidity	Yes	NO	Windy	Yes	NO
Sunny	2	3	hot	2	2	high	3	4	False	6	2
Overcast	4	0	mild	4	2	normal	6	1	True	3	3
Rainy	3	2	cool	3	1						
Total	9	5	Total	9	5	Total	9	5	Total	9	5

Play yes - 9 no - 5

Summarization of count calculations of all input attributes.

Let us now calculate the probability of play being or not for each value of input attribute. Example: 2 Instances for play being not hold when outlook is rainy, and there are in total 5 instances for the outlook attribute where play is not hold.

Probability for play no given outlook rainy.

$$= \frac{\text{Occurrence of no in play class on Outlook = rainy}}{\text{total no. of occurrence for play no for the outlook att}}$$

	Play			Play			Play			Play	
Outlook	Yes	NO	Temp	Yes	NO	Humidity	Yes	NO	Windy	Yes	NO
Sunny	0.22	0.60	hot	0.22	0.40	high	0.33	0.80	false	0.67	0.40
Overcast	0.44	0.00	mild	0.44	0.40	normal	0.67	0.20	true	0.33	0.60
Rainy	0.33	0.40	cool	0.33	0.20						

Play	
Yes	0.64
NO	0.36

Let us predict the output class for some X, let us suppose X is as given below:

Outlook	Temprature	Humidity	Windy	Play
Sunny	Cool	high	true	?

Then, we have to calculate probability of class C_i, when X has already occurred. Thus, it is

$$P(C_i \mid sunny, cool, high, true) = P(sunny, cool, high, true \mid C_i) * P(C_i)$$

- let us first calculate probability of each class i.e. $P(C_i)$
- Here, we have two classes, i.e. Yes and No
- There are total 14 instances in the given dataset from which 9 instances are for class Yes and 5 for class No. Thus

Probability of class Yes, i.e. $P(Yes) = 9/14 = 0.64$
$P(NO) = 5/14 = 0.36$

Now, let us calculate $P(X \mid C_i)$, i.e. $P(sunny, cool, high, true \mid C_i)$

$$P(sunny, cool, high, true \mid C_i) = P(outlook = sunny \mid C_i) * P(Temperature\ cool \mid C_i) * P(humidity\ high \mid C_i) * P(windy = true \mid C_i)$$

$P(outlook = sunny \mid Yes) = 0.22$
$P(Temperature = cool \mid Yes) = 0.33$
$P(Humidity = high \mid Yes) = 0.33$
$P(windy = true \mid Yes) = 0.33$
$P(Yes) = 0.64$

likelihood of $P(X \mid Yes) * P(Yes) = 0.22 * 0.33 * 0.33 * 0.64$
$= 0.0053$

Play being held

Registration No. .... Question ....

Do not write in this margin

Write on both sides of the paper

P(outlook = sunny | NO) = 0.60
P(Temperature = cool | NO) = 0.20
P(humidity = high | NO) = 0.80
P(windy = true | NO) = 0.60
P(NO) = 0.36

= 0.60

Likelihood of play being held = P(X/NO) * P(NO) =

0.60 * 0.20 * 0.80 * 0.60 * 0.36 = 0.0206.

Now, calculated likelihoods are converted to probability to decide whether the play is held or not under these weather conditions'

Probability of play yes = likelihood for yes / likelihood for yes + likelihood for play no).

$$= \frac{0.0053}{(0.0053 + 0.0206)} = 20.5\%$$

Probability of play no = likelihood for play NO / (likelihood for play Yes + likelihood for play NO)

$$= \frac{0.0206}{0.0053 + 0.0206} = 79.5$$

MASENO UNIVERSITY

Registration No. .... Question ....

Write on both sides of the paper

Do not write in this margin

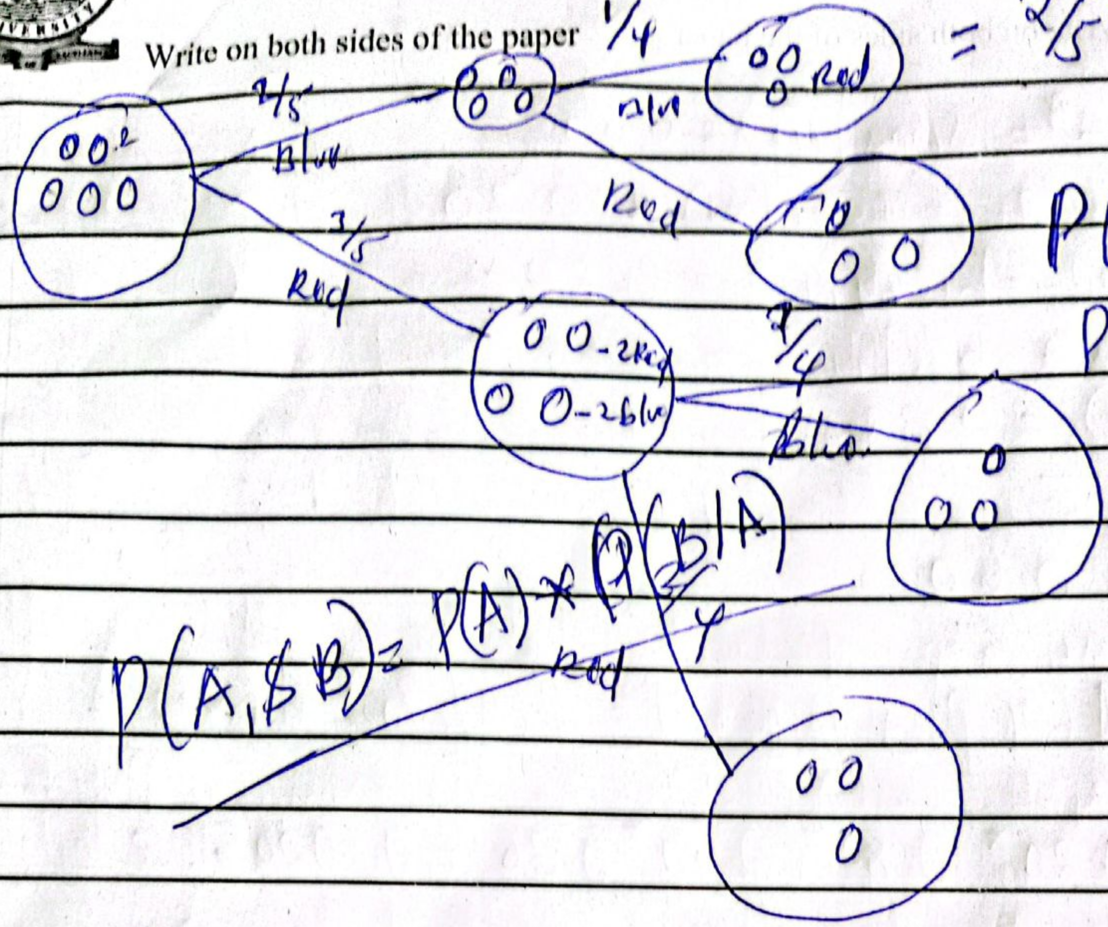
From this we can conclude that there are approximately 80 % chances that play will not be held and 20 % chances that the play will be held.

- This makes a good prediction and Naive Bayes has performed very well in this case.

Registration No. .... Question ....

Do not write In this margin

Write on both sides of the paper



$= 2/5 \times 1/4$

$= 3/20$

$P(A) = \text{Blue}$

$P(B|A) = 1/4$

$P(A, \$ B) = P(A) \times P(B/A)$

$P(A) = 2/5$

$P(B/A) = 1/4$

$P(A \text{ and } B) = P(A) \times P(B/A$

$P(B/A) = \dfrac{P(A \$ B)}{PB}$

$P(A/B) = \dfrac{P(A \cup B)}{P(B)}$

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

$P(B/A) = \dfrac{P(A|B) \times P(B)}{PB}$

$P(B \$ A) = (P(B/A) \times PA)$

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