Machine Learning
ICT-4261

By-
Dr. Jesmin Akhter
Professor
Institute of Information Technology
Jahangirnagar University

Contents

The course will mainly cover the following topics:

- ✔ A Gentle Introduction to Machine Learning
- ✔ Logistic Regression
- ✔ Naive Bayes
- ✔ Support Vector Machines
- ✔ Decision Trees and Ensemble Learning
- ✔ Clustering Fundamentals
- ✔ Hierarchical Clustering
- ✔ Neural Networks and Deep Learning
- ✔ Unsupervised Learning

Outline

- ✔ The Bayesian "Decision"
- ✔ The "Naïve Bayesian" Assumption
- ✔ Types of Naïve Bayes
- ✔ **Laplace Correction**

How Naive Bayes classifier works?

- ✔ Let's understand the working of Naive Bayes through an example.
  - Given an example of *weather conditions and playing sports*.
  - You need to calculate the probability of ***playing sports***.
  - Now, you need to classify whether players will ***play or not***, based on the weather condition.
- ✔ **First Approach (In case of a single feature)**
  - Naive Bayes classifier calculates the probability of an event in the following steps:
    - Step 1: Calculate the prior probability for given class labels
    - Step 2: Find Likelihood probability with each attribute for each class
    - Step 3: Put these value in Bayes Formula and calculate posterior probability.
    - Step 4: See which class has a higher probability, given the input belongs to the higher probability class.

✔ For simplifying prior and posterior probability calculation

- you can use the two tables frequency and likelihood tables.
- Both of these tables will help you to calculate the prior and posterior probability.

- The Frequency table contains the occurrence of labels for all features.
- There are two likelihood tables.
  - Likelihood Table 1 is showing prior probabilities of labels and
  - Likelihood Table 2 is showing the posterior prob

Whether	Play
Sunny	No
Sunny	No
Overcast	Yes
Rainy	Yes
Rainy	Yes
Rainy	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rainy	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table

Whether	No	Yes
Overcast		4
Sunny	2	3
Rainy	3	2
Total	5	9

Likelihood Table 1

Whether	No	Yes		
Overcast		4	=4/14	0.29
Sunny	2	3	=5/14	0.36
Rainy	3	2	=5/14	0.36
Total	5	9		
	=5/14	=9/14		
	0.36	0.64		

Likelihood Table 2

Whether	No	Yes	Posterior Probability for No	Posterior Probability for Yes
Overcast		4	0/5=0	4/9=0.44
Sunny	2	3	2/5=0.4	3/9=0.33
Rainy	3	2	3/5=0.6	2/9=0.22
Total	5	9		

Now suppose you want to calculate the probability of playing when the weather is overcast.

Probability of playing:

P(Yes | Overcast) = P(Overcast | Yes) P(Yes) / P (Overcast) ....(1)

1. Calculate Prior Probabilities:

   P(Overcast) = 4/14 = 0.29

   P(Yes)= 9/14 = 0.64

1. Calculate Posterior Probabilities:

   P(Overcast |Yes) = 4/9 = 0.44

1. Put Prior and Posterior probabilities in equation (1)

   P (Yes | Overcast) = 0.44 * 0.64 / 0.29 = 0.98(Higher)

Similarly, you can calculate the probability of not playing:

Probability of not playing:

P(No | Overcast) = P(Overcast | No) P(No) / P (Overcast) ....(2)

1. Calculate Prior Probabilities:

   P(Overcast) = 4/14 = 0.29

   P(No)= 5/14 = 0.36

1. Calculate Posterior Probabilities:

   P(Overcast |No) = 0/9 = 0

1. Put Prior and Posterior probabilities in equation (2)

   P (No | Overcast) = 0 * 0.36 / 0.29 = 0

The probability of a 'Yes' class is higher. So you can determine here if the weather is overcast than players will play the sport.

Second Approach (In case of multiple features)

Whether	Temperature	Play
Sunny	Hot	No
Sunny	Hot	No
Overcast	Hot	Yes
Rainy	Mild	Yes
Rainy	Cool	Yes
Rainy	Cool	No
Overcast	Cool	Yes
Sunny	Mild	No
Sunny	Cool	Yes
Rainy	Mild	Yes
Sunny	Mild	Yes
Overcast	Mild	Yes
Overcast	Hot	Yes
Rainy	Mild	No

01 CALCULATE PRIOR PROBABILITY FOR GIVEN CLASS LABELS

02 CALCULATE CONDITIONAL PROBABILITY WITH EACH ATTRIBUTE FOR EACH CLASS

03 MULTIPLY SAME CLASS CONDITIONAL PROBABILITY.

04 MULTIPLY PRIOR PROBABILITY WITH STEP 3 PROBABILITY.

05 SEE WHICH CLASS HAS HIGHER PROBABILITY, HIGHER PROBABILITY CLASS BELONGS TO GIVEN INPUT SET STEP.

Now suppose you want to calculate the probability of playing when the weather is overcast, and the temperature is mild.

Probability of playing:

*P(Play= Yes | Weather=Overcast, Temp=Mild) = P(Weather=Overcast, Temp=Mild | Play= Yes)*P(Play=Yes) ....(1)

P(Weather=Overcast, Temp=Mild | Play= Yes)= P(Overcast |Yes) P(Mild |Yes) ....(2)

1. Calculate Prior Probabilities: P(Yes)= 9/14 = 0.64

2. Calculate Posterior Probabilities: P(Overcast |Yes) = 4/9 = 0.44 P(Mild |Yes) = 4/9 = 0.44

3. Put Posterior probabilities in equation (2) P(Weather=Overcast, Temp=Mild | Play= Yes) = 0.44 * 0.44 = 0.1936(Higher)

4. Put Prior and Posterior probabilities in equation (1) P(Play= Yes | Weather=Overcast, Temp=Mild) = 0.1936*0.64 = 0.124

Similarly, you can calculate the probability of not playing:

Probability of not playing:

*P(Play= No | Weather=Overcast, Temp=Mild) = P(Weather=Overcast, Temp=Mild | Play= No)*P(Play=No) ....(3)

P(Weather=Overcast, Temp=Mild | Play= No)= P(Weather=Overcast |Play=No) P(Temp=Mild | Play=No) ....(4)

1. Calculate Prior Probabilities: P(No)= 5/14 = 0.36
2. Calculate Posterior Probabilities: P(Weather=Overcast |Play=No) = 0/9 = 0
   P(Temp=Mild | Play=No)=2/5=0.4
3. Put posterior probabilities in equation (4) P(Weather=Overcast, Temp=Mild | Play= No) = 0 * 0.4= 0
4. Put prior and posterior probabilities in equation (3) P(Play= No | Weather=Overcast, Temp=Mild) = 0*0.36=0

The probability of a 'Yes' class is higher. So you can say here that if the weather is overcast than players will play the sport.

Estimate the probability values for the new instances (Color=green, Legs=2, Height=Tall and Smelly=No)

No	Color	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

I. Calculate prior probability

- ✔ P(M)=4/8= 0.5
- ✔ P(H)= 4/8= 0.5

1. Calculate Conditional probability

Color	M	H
White	2/4	3/4
Green	2/4	1/4

Legs	M	H
2	1/4	4/4
3	3/4	0/4

Height	M	H
Tall	1/4	2/4
Short	3/4	2/4

Smelly	M	H
Yes	3/4	1/4
No	1/4	3/4

Now calculate probability values for the new instances (Color=green, Legs=2, Height=Tall and Smelly=No)

✔ P(**M**|color=green, Legs=2, Height=Tall and Smelly=No)

= *P*(M) * *P*(color=green|M) *

P(Legs=2 |M) *

P(Height=Tall |M) *

P(Smelly=No |M)

= 0.5 * 0.5 * .25 * .25 * .25

=1/256

=0.0039

Color	M	H
White	2/4	3/4
Green	2/4	1/4

Legs	M	H
2	1/4	4/4
3	3/4	0/4

Height	M	H
Tall	3/4	2/4
Short	1/4	2/4

Smelly	M	H
Yes	3/4	1/4
No	1/4	3/4

✔ P(**H**|color=green, Legs=2, Height=Tall and Smelly=No)

= *P*(H) * *P*(color=green|H) *

P(Legs=2 |H) *

P(Height=Tall |H) *

P(Smelly=No |H)

= 0.5 * 0.25 * 1 * .5 * ¾

=3/64

=0.047

Color	M	H
White	2/4	3/4
Green	2/4	1/4

Legs	M	H
2	1/4	4/4
3	3/4	0/4

Height	M	H
Tall	3/4	2/4
Short	1/4	2/4

Smelly	M	H
Yes	3/4	1/4
No	1/4	3/4

Since p(H|new instance)> p(M|new instance)

So the new instance belongs to Species **H**

age	income	student	Credit rating	Buys computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

- **New Example**
- E → age<=30, income=medium, student=yes, credit-rating=fair
- Where,
- E_1 is age<=30,
- E2 is income=medium,
- E_3 student=yes,
- E_4 is credit-rating=fair
- We need to compute **P(yes|E)** and **P(no|E)** and compare them.

age	income	student	Credit rating	Buys computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

- E_1 is age<=30,
- E2 is income=medium,
- E_3 student=yes,
- E_4 is credit-rating=fair
- We need to compute **P(yes|E)** and **P(no|E)** and compare them.

$$P(\, yes|E) = \frac{P(E_1|yes)\ P(E_2|yes)\ P(E_3|yes)\ P(E_4|yes)\ P(yes)}{P(E)}$$

$$P(\, no|E) = \frac{P(E_1|no)\ P(E_2|no)\ P(E_3|no)\ P(E_4|no)\ P(no)}{P(E)}$$

age	income	student	Credit rating	Buys computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

$$P(yes|E) = \frac{P(E_1|yes)\,P(E_2|yes)\,P(E_3|yes)\,P(E_4|yes)\,P(yes)}{P(E)}$$

$$P(no|E) = \frac{P(E_1|no)\,P(E_2|no)\,P(E_3|no)\,P(E_4|no)\,P(no)}{P(E)}$$

$P(yes)=9/14=0.643$ **$P(no)=5/14=0.357$**

$P(E_1|yes)=2/9=0.222$ **$P(E1|no)=3/5=0.6$**

$P(E_2|yes)=4/9=0.444$ **$P(E2|no)=2/5=0.4$**

$P(E_3|yes)=6/9=0.667$ **$P(E3|no)=1/5=0.2$**

$P(E_4|yes)=6/9=0.667$ **$P(E4|no)=2/5=0.4$**

E_1 is age<=30,

E2 is income=medium

E_3 student=yes,

E_4 is credit-rating=fair

age	income	student	Credit rating	Buys computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

$$P(yes|E) = \frac{0.222 * 0.444 * 0.667 * 0.667 * 0.643}{P(E)} = \frac{0.028}{P(E)}$$

$$P(no|E) = \frac{0.6 * 0.4 * 0.2 * 0.4 * 0.357}{P(E)} = \frac{0.007}{P(E)}$$

Hence, the Naïve Bayes classifier predicts

buys_computer = yes for the new example.

Naive assumption
Example 3

- ✔ Let us try to apply the Naive formula manually on our weather dataset.
- ✔ For example, probability of playing golf given that the temperature is cool, i.e P(temp. = cool | play golf = Yes) = 3/9.
- ✔ Also, we need to find class probabilities (P(y)) which has been calculated in the table 5. For example, P(play golf = Yes) = 9/14.

	Yes	No	P(Yes)	P(No)
Sunny	3	2	3/9	2/5
Overcast	4	0	4/9	0/5
Rainy	2	3	2/9	3/5
Total	9	5	100%	100%

Outlook

	Yes	No	P(yes)	P(no)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rainy	3	2	3/9	2/5
Total	9	5	100%	100%

Temperature

	Yes	No	P(yes)	P(no)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Total	9	5	100%	100%

Humidity

	Yes	No	P(yes)	P(no)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
Total	9	5	100%	100%

Wind

	Yes	No	P(yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
Total	9	5	100%	100%

Play		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
Total	14	100%

Naive assumption

- ✔ So now, we are done with our pre-computations and the classifier is ready!
- ✔ Let us test it on a new set of features (let us call it today):
- ✔ today = (Sunny, Hot, Normal, False)
- ✔ So, probability of playing golf is given by:

$$P(Yes|today) = \frac{P(SunnyOutlook|Yes)P(HotTemperature|Yes)P(NormalHumidity|Yes)P(NoWind|Yes)P(Yes)}{P(today)}$$

- ✔ and probability to not play golf is given by:

$$P(No|today) = \frac{P(SunnyOutlook|No)P(HotTemperature|No)P(NormalHumidity|No)P(NoWind|No)P(No)}{P(today)}$$

Naive assumption

✔ Since, P(today) is common in both probabilities, we can ignore P(today) and find proportional probabilities as:

$$P(Yes|today) \propto \frac{2}{9} \cdot \frac{2}{9} \cdot \frac{6}{9} \cdot \frac{6}{9} \cdot \frac{9}{14} \approx 0.0141$$

and

$$P(No|today) \propto \frac{3}{5} \cdot \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{2}{5} \cdot \frac{5}{14} \approx 0.0068$$

Now, since

$$P(Yes|today) + P(No|today) = 1$$

Naive assumption

- ✔ These numbers can be converted into a probability by making the sum equal to 1 (normalization):

$$P(Yes|today) = \frac{0.0141}{0.0141+0.0068} = 0.67$$

and

$$P(No|today) = \frac{0.0068}{0.0141+0.0068} = 0.33$$

Since

$$P(Yes|today) > P(No|today)$$

- ✔ So, prediction that golf would be played is ‘Yes’.
- ✔ The method that we discussed above is applicable for discrete data. In case of continuous data, we need to make some assumptions regarding the distribution of values of each feature. The different naive Bayes classifiers differ mainly by the assumptions they make regarding the distribution of $P(x_i \mid y)$.

Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$\langle Outlook = sunny, Temperature = cool, Humidity = high, Wind = strong \rangle$$

$$P(PlayTennis = yes) = 9/14 = .64$$

$$P(PlayTennis = no) = 5/14 = .36$$

Outlook	Y	N
sunny	2/9	3/5
overcast	4/9	0
rain	3/9	2/5
Tempreature		
hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Humidity	Y	N
high	3/9	4/5
normal	6/9	1/5
Windy		
Strong	3/9	3/5
Weak	6/9	2/5

$$\langle Outlook = sunny, Temperature = cool, Humidity = high, Wind = strong \rangle$$

$$v_{NB} = \underset{v_j \in \{yes, no\}}{\text{argmax}} \ P(v_j) \prod_i P(a_i|v_j)$$

$$= \underset{v_j \in \{yes, no\}}{\text{argmax}} \ P(v_j) \quad P(Outlook = sunny|v_j) P(Temperature = cool|v_j)$$

$$\cdot \ P(Humidity = high|v_j) P(Wind = strong|v_j)$$

$$v_{NB}(yes) = P(yes) \ P(sunny|yes) \ P(cool|yes) \ P(high|yes) \ P(strong|yes) = .0053$$

$$v_{NB}(no) = P(no) \ P(sunny|no) \ P(cool|no) \ P(high|no) \ P(strong|no) = .0206$$

$$v_{NB}(yes) = \frac{v_{NB}(yes)}{v_{NB}(yes) + v_{NB}(no)} = 0.205 \qquad v_{NB}(no) = \frac{v_{NB}(no)}{v_{NB}(yes) + v_{NB}(no)} = 0.795$$

which has higher probability.
So will not go to play

- ✔ **Problem**: If the weather is sunny, then the Player should play or not?
- ✔ **Solution**: To solve this, first consider the below dataset:

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

- ✔ **Frequency table for the Weather Conditions:**

Weather	Yes	No
Overcast	5	0
Rainy	2	2
Sunny	3	2
Total	10	4

- ✔ **Likelihood table weather condition:**

Weather	No	Yes	
Overcast	0	5	5/14= 0.35
Rainy	2	2	4/14=0.29
Sunny	2	3	5/14=0.35
All	4/14=0.29	10/14=0.71	

- **Applying Bayes'theorem:**
- **P(Yes|Sunny)= P(Sunny|Yes)*P(Yes)/P(Sunny)**
- P(Sunny|Yes)= 3/10= 0.3
- P(Sunny)= 0.35
- P(Yes)=0.71
- So P(Yes|Sunny) = 0.3*0.71/0.35= **0.60**
- **P(No|Sunny)= P(Sunny|No)*P(No)/P(Sunny)**
- P(Sunny|NO)= 2/4=0.5
- P(No)= 0.29
- P(Sunny)= 0.35
- So P(No|Sunny)= 0.5*0.29/0.35 = **0.41**
- So as we can see from the above calculation that **P(Yes|Sunny)>P(No|Sunny)**
- **Hence on a Sunny day, Player can play the game.**

- ✔ Let us take an example to get some better intuition. Consider the car theft problem with attributes Color, Type, Origin, and the target, Stolen can be either Yes or No. **we need to classify whether the car is stolen, given the features of the car**.
- ✔ The columns represent these features and the rows represent individual entries. If we take the first row of the dataset, we can observe that the car is stolen if the Color is Red, the Type is Sports and Origin is Domestic. So **we want to classify a Red Domestic SUV is getting stolen or not.** Note that there is no example of a **Red Domestic SUV** in our data set.

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Frequency Table		Stolen?	
		Yes	No
Origin	Domestic	2	3
	Imported	3	2

Likelihood Table		Stolen?	
		P(Yes)	P(No)
Origin	Domestic	2/5	3/5
	Imported	3/5	2/5

Frequency and Likelihood tables of 'Origin'

So in our example, we have 3 predictors **X**.

Color	Type	Origin	Stolen
Red	SUV	Domestic	?

Correct the error

- ✔ Since 0.144 > 0.048, Which means given the features RED SUV and Domestic, our example gets classified as 'NO' the car is not stolen.

P(Yes | X) = P(Red | Yes) * P(SUV | Yes) * P(Domestic | Yes) * P(Yes)

$= \frac{3}{5} * \frac{1}{5} * \frac{2}{5} * 1$

= 0.048

P(No | X) = P(Red | No) * P(SUV | No) * P(Domestic | No) * P(No)

$= \frac{2}{5} * \frac{3}{5} * \frac{3}{5} * 1$

= 0.144

Example of Naïve Bayes Classifier

[Optional]

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

A: attributes

M: mammals

N: non-mammals

$$P(A|M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} \times \frac{2}{7} = 0.06$$

$$P(A|N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042$$

$$P(A|M)P(M) = 0.06 \times \frac{7}{20} = 0.021$$

$$P(A|N)P(N) = 0.004 \times \frac{13}{20} = 0.0027$$

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

P(A|M)P(M) > P(A|N)P(N)

=> Mammals

- ✔ For the sake of computing the probabilities, let's aggregate the training data to form a counts table like this.

Type	Long	Not Long	Sweet	Not Sweet	Yellow	Not Yellow	**Total**
Banana	400	100	350	150	450	50	500
Orange	0	300	150	150	300	0	300
Other	100	100	150	50	50	150	200
Total	500	500	650	350	800	200	1000

- ✔ So the objective of the classifier is to predict if a given fruit is a 'Banana' or 'Orange' or 'Other' when only the 3 features (long, sweet and yellow) are known.
- ✔ Let's say you are given a fruit that is: Long, Sweet and Yellow, can you predict what fruit it is?

- ✔ **Step 1: Compute the 'Prior' probabilities for each of the class of fruits.**

Out of 1000 records in training data, you have 500 Bananas, 300 Oranges and 200 Others.

So the respective priors are

P(Y=Banana) = 500 / 1000 = 0.50

P(Y=Orange) = 300 / 1000 = 0.30

P(Y=Other) = 200 / 1000 = 0.20

- ✔ **Step 2: Compute the probability of evidence that goes in the denominator.**

This is nothing but the product of P of Xs for all X. This is an optional step because the denominator is the same for all the classes and so will not affect the probabilities.

P(x1=Long) = 500 / 1000 = 0.50

P(x2=Sweet) = 650 / 1000 = 0.65

P(x3=Yellow) = 800 / 1000 = 0.80

✔ **<u>Step 3: Compute the probability of likelihood of evidences that goes in the numerator.</u>**

In the table, you have 500 Bananas. Out of that 400 is long.

Probability of Likelihood for Banana

P(x1=Long | Y=Banana) = 400 / 500 = 0.80

P(x2=Sweet | Y=Banana) = 350 / 500 = 0.70

P(x3=Yellow | Y=Banana) = 450 / 500 = 0.90.

So, the overall probability of Likelihood of evidence for Banana = 0.8 * 0.7 * 0.9 = 0.504

- ✔ **<u>Step 4: Substitute all the 3 equations into the Naive Bayes formula, to get the probability that it is a banana.</u>**

<u>Step 4: If a fruit is *'Long', 'Sweet' and 'Yellow'*, what fruit is it?</u>

$$P(\text{Banana} \mid \textit{Long, Sweet and Yellow}) = \frac{P(\text{Long} \mid \text{Banana}) * P(\text{Sweet} \mid \text{Banana}) * P(\text{Yellow} \mid \text{Banana}) \times P(\text{banana})}{P(\text{Long}) * P(\text{Sweet}) * P(\text{Yellow})}$$

$$= \frac{0.8 * 0.7 * 0.9 * 0.5}{P(\text{Evidence})} = 0.252/P(\text{Evidence})$$

$P(\text{Orange} \mid \textit{Long, Sweet and Yellow}) =$ 0, because $P(\text{Long} \mid \text{Orange}) = 0$

$P(\text{Other Fruit} \mid \textit{Long, Sweet and Yellow}) = 0.01875 / P(\text{Evidence})$

Answer: Banana - Since it has highest probability amongst the 3 classes

- ✔ Similarly, you can compute the probabilities for 'Orange' and 'Other fruit'. The denominator is the same for all 3 cases, so it's optional to compute.
- ✔ Clearly, Banana gets the highest probability, so that will be our predicted class.

Types of Naïve Bayes

There are three main types of Naïve Bayes which are listed below:

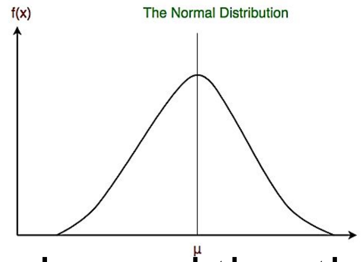
1. Gaussian Naïve Bayes
2. Multinomial Naïve Bayes
3. Bernoulli Naïve Bayes

✔ In addition

- **Complement Naive Bayes**
- **Categorical Naive Bayes**

Gaussian Naive Bayes classifier

- ✔ In Gaussian Naive Bayes, continuous values associated with each feature are assumed to be distributed according to a **Gaussian distribution**. Gaussian distribution is also called normal distribution. Normal distribution is a statistical model that describes the distributions of continuous random variables in nature and is defined by its bell-shaped curve.
- ✔ The two most important features of the normal distribution are the mean (μ) and standard deviation (σ). The mean is the average value of a distribution, and the standard deviation is the "width" of the distribution around the mean.
- ✔ A variable (X) that is normally distributed is distributed continuously (continuous variable) from $-\infty < X < +\infty$, and the total area under the model curve is 1.



- ✔ The attribute is first segmented based on the output class, and then the variance and mean of the attribute are calculated for each class.
- ✔ The combination of the prediction for all parameters is the final prediction that returns a probability of the dependent variable to be classified in each group. The final classification is assigned to the group with the higher probability.

Gaussian Naive Bayes classifier

- ✔ Gaussian naive Bayes is useful when working with continuous values whose probabilities can be modeled using a Gaussian distribution

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- ✔ In the above formulae, sigma and mu is the variance and mean of the continuous variable X computed for a given class of Y.

Gaussian Naive Bayes Algorithm – Solved Example 6

Person	Height (ft)	Weight (lbs)	Foot size (inches)
Male	6.00	180	12
Male	5.92	190	11
Male	5.58	170	12
Male	5.92	165	10
Female	5.00	100	6
Female	5.50	150	8
Female	5.42	130	7
Female	5.75	150	9

Based on the following data determine the gender of a person having height 6 ft., weight 130 lbs, and foot size 8 inch. (use Naive Bayes algorithm).

Gaussian Naive Bayes Algorithm – Solved Example 6

Person	Height (ft)	Weight (lbs)	Foot size (inches)
Male	6.00	180	12
Male	5.92	190	11
Male	5.58	170	12
Male	5.92	165	10
Female	5.00	100	6
Female	5.50	150	8
Female	5.42	130	7
Female	5.75	150	9

$P(Male) = 4/8 = 0.5$

$P(Female) = 4/8 = 0.5$

Male:

$$Mean\ (Height) = \frac{(6+5.92+5.58+5.92)}{4} = 5.855$$

$$Variance\ (Height) = \frac{\sum(x_i - \bar{x})^2}{n-1} \quad Variance = (SD)^2 = \sigma^2 = \sum\frac{(x-\bar{x})^2}{n-1}$$

$$= \frac{(6-5.855)^2+(5.92-5.855)^2+(5.58-5.855)^2+(5.92-5.855)^2}{4-1}$$

$=$ **0.035033**

Gaussian Naive Bayes Algorithm – Solved Example 6

Sex	Mean (height)	Variance (height)	Mean (weight)	Variance (weight)	Mean(foot size)	Variance (foot size)
Male	5.855	0.035033	176.25	122.92	11.25	0.91667
Female	5.4175	0.097225	132.5	0558.33	7.5	1.6667

New Instance to be Classified is:

Sex	Height(ft)	Weight(lbs)	Foot size(inch)
Sample	6	130	8

$$P(Male) = 4/8 = 0.5$$
$$P(Female) = 4/8 = 0.5$$

Person	Height (ft)	Weight (lbs)	Foot size (inches)
Male	6.00	180	12
Male	5.92	190	11
Male	5.58	170	12
Male	5.92	165	10
Female	5.00	100	6
Female	5.50	150	8
Female	5.42	130	7
Female	5.75	150	9

$$Posterior\ (Male) = \frac{P(M) * P(H|M) * P(W|M) * P(FS|M)}{Evidence}$$

$$Posterior\ (Female) = \frac{P(F) * P(H|F) * P(W|F) * P(FS|F)}{Evidence}$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$P(H|M) = \frac{1}{\sqrt{2 * 3.142 * 0.035033}} * e^{-\frac{(6-5.855)^2}{2 * 0.035033}} = 1.5789$$

$$P(W|M) = 5.9881e^{-6}$$

$$P(FS|M) = 1.3112e^{-3}$$

$$P(H|F) = 2.2346e^{-1}$$

$$P(W|F) = 1.6789e^{-2}$$

$$P(FS|F) = 2.8669e^{-1}$$

$$Posterior\ (Male) = \frac{P(M) \cdot P(H|M) \cdot P(W|M) \cdot P(FS|M)}{Evidence} = 0.5 * 1.5789 * 5.9881e^{-6} * 1.3112e^{-3} = 6.1984e^{-9}$$

$$Posterior\ (Female) = \frac{P(F) \cdot P(H|F) \cdot P(W|F) \cdot P(FS|F)}{Evidence} = 0.5 * 2.2346e^{-1} * 1.6789e^{-2} * 2.8669e^{-1} = 5.377e^{-4}$$

Female

Laplace Smoothing

- ✔ If the value of P(Long | Orange) is zero, that is, there is no 'Long' oranges in the training data then this will make P(Orange | Long, Sweet and Yellow) equal to zero. **This is the problem of zero probability.**
- ✔ **So, how to deal with this problem?**
  - Laplace smoothing is a smoothing technique that handles the problem of zero probability in Naïve Bayes
  - To avoid this, we increase the count of the variable with zero to a small value (usually 1) in the numerator, so that the overall probability doesn't become zero. This approach is called **'Laplace Correction' or Laplace Smoothing**
  - Now we will solve the following zero probability problem using **Laplace Smoothing** technique.
  - .

$$P(\text{Overcast} \mid \text{No}) = 0/9 = 0$$

$$P(\text{No} \mid \text{Overcast}) = 0 * 0.36 / 0.29 = 0$$

Gaussian Naive Bayes Classifier – Solved Example 1

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	85	85	False	No
D2	Sunny	80	90	True	No
D3	Overcast	83	86	False	Yes
D4	Rainy	70	96	False	Yes
D5	Rainy	68	80	False	Yes
D6	Rainy	65	70	True	No
D7	Overcast	64	65	True	Yes
D8	Sunny	72	95	False	No
D9	Sunny	69	70	False	Yes
D10	Rainy	75	80	False	Yes
D11	Sunny	75	70	True	Yes
D12	Overcast	72	90	True	Yes
D13	Overcast	81	75	False	Yes
D14	Rainy	71	91	True	No

- First, we can calculate the probabilities for the nominal attributes:
- $P(yes) = \frac{9}{14} = 0.643$
- $P(no) = \frac{5}{14} = 0.357$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	85	85	False	No
D2	Sunny	80	90	True	No
D3	Overcast	83	86	False	Yes
D4	Rainy	70	96	False	Yes
D5	Rainy	68	80	False	Yes
D6	Rainy	65	70	True	No
D7	Overcast	64	65	True	Yes
D8	Sunny	72	95	False	No
D9	Sunny	69	70	False	Yes
D10	Rainy	75	80	False	Yes
D11	Sunny	75	70	True	Yes
D12	Overcast	72	90	True	Yes
D13	Overcast	81	75	False	Yes
D14	Rainy	71	91	True	No

- $outlook = overcast, temperature = 60, humidity = 62, windy = false.$
- $P(outlook = overcast|yes) = \frac{4}{9} = 0.444$
- $P(outlook = overcast|no) = \frac{0}{5} = 0$
- $P(windy = false|yes) = \frac{6}{9} = 0.667$
- $P(windy = false|no) = \frac{2}{5} = 0.4$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	85	85	False	No
D2	Sunny	80	90	True	No
D3	Overcast	83	86	False	Yes
D4	Rainy	70	96	False	Yes
D5	Rainy	68	80	False	Yes
D6	Rainy	65	70	True	No
D7	Overcast	64	65	True	Yes
D8	Sunny	72	95	False	No
D9	Sunny	69	70	False	Yes
D10	Rainy	75	80	False	Yes
D11	Sunny	75	70	True	Yes
D12	Overcast	72	90	True	Yes
D13	Overcast	81	75	False	Yes
D14	Rainy	71	91	True	No

- As **P(outlook=overcast|no)=0**, we need to use a Laplace estimator for the attribute outlook.
- We assume that the three values (sunny, overcast, rainy) are equally probable and set X=3 and k = 1:

$$P_{LAP,k}(x|y) = \frac{c(x,y)+k}{c(y)+k|X|}$$

- $P(outlook = overcast|yes) = \frac{4+1}{9+3} = 0.4167$
- $P(outlook = overcast \mid no) = \frac{0+1}{5+3} = 0.125$

- Second, we need to calculate the mean μ and standard deviation σ values for the numerical attributes.
- $X_i, i = 1..n$, n-number of measurements

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	85	85	False	No
D2	Sunny	80	90	True	No
D3	Overcast	83	86	False	Yes
D4	Rainy	70	96	False	Yes
D5	Rainy	68	80	False	Yes
D6	Rainy	65	70	True	No
D7	Overcast	64	65	True	Yes
D8	Sunny	72	95	False	No
D9	Sunny	69	70	False	Yes
D10	Rainy	75	80	False	Yes
D11	Sunny	75	70	True	Yes
D12	Overcast	72	90	True	Yes
D13	Overcast	81	75	False	Yes
D14	Rainy	71	91	True	No

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	85	85	False	No
D2	Sunny	80	90	True	No
D3	Overcast	83	86	False	Yes
D4	Rainy	70	96	False	Yes
D5	Rainy	68	80	False	Yes
D6	Rainy	65	70	True	No
D7	Overcast	64	65	True	Yes
D8	Sunny	72	95	False	No
D9	Sunny	69	70	False	Yes
D10	Rainy	75	80	False	Yes
D11	Sunny	75	70	True	Yes
D12	Overcast	72	90	True	Yes
D13	Overcast	81	75	False	Yes
D14	Rainy	71	91	True	No

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \sigma = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n-1}$$

$$\mu(Temp_{yes}) = \frac{(83 + 70 + \cdots + 81)}{9} = 73$$

$$\sigma(Temp_{yes})$$

$$= \frac{(83-73)^2 + (70-73)^2 \ldots + (81-73)^2}{9-1}$$

$$= 6.2$$

		μ	σ
Temp	Yes	73	6.2
	No	74.6	8.0
Hum	Yes	79.1	10.2
	No	86.2	9.7

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	85	85	False	No
D2	Sunny	80	90	True	No
D3	Overcast	83	86	False	Yes
D4	Rainy	70	96	False	Yes
D5	Rainy	68	80	False	Yes
D6	Rainy	65	70	True	No
D7	Overcast	64	65	True	Yes
D8	Sunny	72	95	False	No
D9	Sunny	69	70	False	Yes
D10	Rainy	75	80	False	Yes
D11	Sunny	75	70	True	Yes
D12	Overcast	72	90	True	Yes
D13	Overcast	81	75	False	Yes
D14	Rainy	71	91	True	No

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	85	85	False	No
D2	Sunny	80	90	True	No
D3	Overcast	83	86	False	Yes
D4	Rainy	70	96	False	Yes
D5	Rainy	68	80	False	Yes
D6	Rainy	65	70	True	No
D7	Overcast	64	65	True	Yes
D8	Sunny	72	95	False	No
D9	Sunny	69	70	False	Yes
D10	Rainy	75	80	False	Yes
D11	Sunny	75	70	True	Yes
D12	Overcast	72	90	True	Yes
D13	Overcast	81	75	False	Yes
D14	Rainy	71	91	True	No

$outlook = overcast, temperature = 60, humidity = 62, windy = false.$

Third, calculate

$f(temperature = 60|yes)$,

$f(temperature = 60|no)$,

$f(humidity = 62|yes)$

$f(humidity = 62|no)$

using the probability density function for the normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$outlook = overcast, temperature = 60, humidity = 62, windy = false.$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	85	85	False	No
D2	Sunny	80	90	True	No
D3	Overcast	83	86	False	Yes
D4	Rainy	70	96	False	Yes
D5	Rainy	68	80	False	Yes
D6	Rainy	65	70	True	No
D7	Overcast	64	65	True	Yes
D8	Sunny	72	95	False	No
D9	Sunny	69	70	False	Yes
D10	Rainy	75	80	False	Yes
D11	Sunny	75	70	True	Yes
D12	Overcast	72	90	True	Yes
D13	Overcast	81	75	False	Yes
D14	Rainy	71	91	True	No

		μ	σ
Temp	Yes	73	6.2
	No	74.6	8.0
Hum	Yes	79.1	10.2
	No	86.2	9.7

the normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$f(temperature = 60 \mid yes) = \frac{1}{6.2\sqrt{2\pi}} e^{-\frac{(60-73)^2}{2\ (6.2)^2}} = 0.007$$

$$f(temperature = 60 \mid no) = \frac{1}{8\sqrt{2\pi}} e^{-\frac{(60-74.6)^2}{2\ 8^2}} = 0.0094$$

$$f(humidity = 62 \mid yes) = \frac{1}{10.2\sqrt{2\pi}} e^{-\frac{(62-79.1)^2}{2\ (10.2)^2}} = 0.0096$$

$$f(humidity = 62 \mid no) = \frac{1}{9.7\sqrt{2\pi}} e^{-\frac{(62-86.2)^2}{2\ (9.7)^2}} = 0.0018$$

- Fourth, we can calculate the final probabilities:
- $\boldsymbol{outlook = overcast, temperature = 60, humidity = 62, windy = false.}$
- $p(yes|E) =$

$$\frac{p(outlook = overcast|yes) * p(temp = 60|yes) * p(hum = 62|yes) * p(windy = false|yes) * p(yes)}{p(E)}$$

$$= \frac{0.4167 * 0.0071 * 0.0096 * 0.667 * 0.643}{p(E)} = \frac{1.22 * 10^{-5}}{p(E)}$$

- $p(no|E) =$

$$\frac{p(outlook = overcast|no) * p(temp = 60|no) * p(hum = 62|no) * p(windy = false|no) * p(no)}{p(E)}$$

$$= \frac{0.125 * 0.0094 * 0.0018 * 0.4 * 0.357}{p(E)} = \frac{3.02 * 10^{-7}}{p(E)}$$

- Therefore, the Naïve Bayes classifier predicts **play=yes** for the new example.

Multnomial Naive Bayes: Example

	docID	words in document	in c = China?
Training set	1	Chinese Beijing Chinese	yes
	2	Chinese Chinese Shanghai	yes
	3	Chinese Macao	yes
	4	Tokyo Japan Chinese	no
Test set	5	Chinese Chinese Chinese Tokyo Japan	?

Feature vectors represent the frequencies with which certain events have been generated by a **multinomial distribution**. This is the event model typically used for document classification/text classification.

$$\hat{\theta}_{yi} = \frac{N_{yi} + \alpha}{N_y + \alpha n}$$

$$P(c) = \frac{3}{4} \qquad P(\bar{c}) = \frac{1}{4}$$

$$P(\text{Chinese}|c) = \frac{(5+1)}{(8+6)} = \frac{6}{14} = \frac{3}{7} \qquad P(\text{Toyko}|c) = P(\text{Japan}|c) = \frac{(0+1)}{(8+6)} = \frac{1}{14}$$

$$P(\text{Chinese}|\bar{c}) = \frac{(1+1)}{(3+6)} = \frac{2}{9} \qquad P(\text{Toyko}|\bar{c}) = P(\text{Japan}|\bar{c}) = \frac{(1+1)}{(3+6)} = \frac{2}{9}$$

$$P(c|d_5) \propto \frac{3}{4} \cdot \left(\frac{3}{7}\right)^3 \cdot \frac{1}{14} \cdot \frac{1}{14} \approx 0.0003$$

$$P(\bar{c}|d_5) \propto \frac{1}{4} \cdot \left(\frac{2}{9}\right)^3 \cdot \frac{2}{9} \cdot \frac{2}{9} \approx 0.0001$$

The classifier assigns the test document to c = China

where n is the number of features (in text classification, the size of the vocabulary) and θ_{yi} is the probability $P(x_i \mid y)$ of feature i appearing in a sample belonging to class y.

where $N_{yi} = \sum_{x \in T} x_i$ is the number of times feature i appears in a sample of class y in the training set T, and $N_y = \sum_{i=1}^{n} N_{yi}$ is the total count of all features for class y.

The smoothing priors $\alpha \geq 0$ accounts for features not present in the learning samples and prevents zero probabilities in further computations. Setting $\alpha = 1$ is called Laplace smoothing, while $\alpha < 1$ is called Lidstone smoothing.

Multinomial Naive Bayes

We have an existing set of documents (D1-D5) that are categorized into Auto, Sports, and Computer.

Document #	Content	Category
D1	Saturn Dealer's Car	Auto
D2	Toyota Car Tercel	Auto
D3	Baseball Game Play	Sports
D4	Pulled Muscle Game	Sports
D5	Colored GIFs Root	Computer

Now the task is to categorize the new D6 and D7 into Auto, Sports, or Computer.

Document #	Content	Category
D6	Home Runs Game	?
D7	Car Engine Noises	?

Step 1

Calculate prior probabilities. These are the probability of a document being in a specific category from the given set of documents.

P(Category) = (No. of documents classified into the category) divided by (Total number of documents)

P(Auto) = (No of documents classified into Auto)divided by (Total number of documents) = $^{2}/_{5}$ = 0.4

P(Sports) = $^{2}/_{5}$ = 0.4

P(Computer) = $^{1}/_{5}$ = 0.2

Step 2

Calculate Likelihood. Likelihood is the conditional probability of a word occurring in a document given that the document belongs to a particular category.

P(Word/Category) = (Number of occurrence of the word in all the documents from a category+1) divided by (All the words in every document from a category + Total number of unique words in all the documents)

P(Saturn/Auto) = (Number of occurrence of the word "SATURN" in all the documents in "AUTO"+1) divided by (All the words in every document from "AUTO" + Total number of unique words in all the documents)

= (1+1)/(6+13) = 2/19 = 0.105263158

Auto

Word	# of Occurrences of Word in Auto	Total Words in Auto	Conditional Probability of Given Word in Auto	# of Total Unique Words in All Documents
Saturn	1	6	0.105263158	13
Dealers	1	6	0.105263158	13
Car	2	6	0.157894737	13
Toyota	1	6	0.105263158	13
Tercel	1	6	0.105263158	13
Baseball	0	6	0.052631579	13
Game	0	6	0.052631579	13
Play	0	6	0.052631579	13
Pulled	0	6	0.052631579	13
Muscle	0	6	0.052631579	13
Colored	0	6	0.052631579	13
GIFs	0	6	0.052631579	13
Root	0	6	0.052631579	13
Home	0	6	0.052631579	13
Runs	0	6	0.052631579	13
Engine	0	6	0.052631579	13
Noises	0	6	0.052631579	13

Sports

Word	# of Occurrences of Word in Sports	Total Words in Sports	Conditional Probability of Given Word	# of Total Unique Words in AI Documents
Saturn	0	6	0.052631579	13
Dealers	0	6	0.052631579	13
Car	0	6	0.052631579	13
Toyota	0	6	0.052631579	13
Tercel	0	6	0.052631579	13
Baseball	1	6	0.105263158	13
Game	2	6	0.157894737	13
Play	1	6	0.105263158	13
Pulled	1	6	0.105263158	13
Muscle	1	6	0.105263158	13
Colored	1	6	0.105263158	13
GIFs	1	6	0.105263158	13
Root	1	6	0.105263158	13
Home	0	6	0.052631579	13
Runs	0	6	0.052631579	13
Engine	0	6	0.052631579	13
Noises	0	6	0.052631579	13

Computer

Word	# of Occurrences of Word in Computer	Total Words in Computer	Conditional Probability of Given Word in Computer	# of Total Unique Words in All Documents
Saturn	0	3	0.0625	13
Dealers	0	3	0.0625	13
Car	0	3	0.0625	13
Toyota	0	3	0.0625	13
Tercel	0	3	0.0625	13
Baseball	0	3	0.0625	13
Game	0	3	0.0625	13
Play	0	3	0.0625	13
Pulled	0	3	0.0625	13
Muscle	0	3	0.0625	13
Colored	1	3	0.125	13
GIFs	1	3	0.125	13
Root	1	3	0.125	13
Home	0	3	0.0625	13
Runs	0	3	0.0625	13
Engine	0	3	0.0625	13
Noises	0	3	0.0625	13

Step 3

Calculate P(Category/Document) = P(Category) * P(Word1/Category) * P(Word2/Category) * P(Word3/Category)

P(Auto/D6) = P(Auto) * P(Home/Auto) * P(Runs/Auto) * P(Game/Auto)

= (0.4) * (0.052631579) * (0.157894737)

= (0.00005831754)

P(Sports/D6) = 0.000174953

P(Computers/D6) = 0.00004882813

The most probable category for D6 to fall into is Sports, because it has the highest probability among its peers.

P(Auto/D7) = 0.00017495262

P(Sports/D7) = 0.0000583175

P(Computers/D7) = 0.00004882813

The most probable category for D7 to fall into is Auto, because it has the highest probability among its peers.

The Multinomial Naive Bayes technique is pretty effective for document classification.

Multinomial Naive Bayes

- ✔ ***P(positive | overall liked the movie)*** *= P(overall liked the movie | positive) * P(positive) / P(overall liked the movie)*

Text	Reviews
“I liked the movie”	positive
“It’s a good movie. Nice story”	positive
“Nice songs. But sadly boring ending. ”	negative
“Hero’s acting is bad but heroine looks good. Overall nice movie”	positive
“Sad, boring movie”	negative

Multinomial Naive Bayes

✔ **Calculating probabilities:**
First, we calculate the a priori probability of each tag: for a given sentence in our training data, the probability that it is positive P(positive) is 3/5. Then, P(negative) is 2/5.
Then, calculating P(overall | positive) means counting how many times the word "overall" appears in positive texts (1) divided by the total number of words in positive (17). Therefore, P(overall | positive) = 1/17, P(liked/positive) = 1/17, P(the/positive) = 2/17, P(movie/positive) = 3/17.
If probability comes out to be zero then By using Laplace smoothing: we add 1 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 total possible words count are 21.

Multinomial Naive Bayes

✔ Applying smoothing, The results are:

Word	P(word | positive)	P(word | negative)
overall	(1 + 1)/(17 + 21)	(0 + 1)/(7 + 21)
liked	(1 + 1)/(17 + 21)	(0 + 1)/(7 + 21)
the	(2 + 1/17 + 21)	(0 + 1)/(7 + 21)
movie	(3 + 1)/(17 + 21)	(1 + 1)/(7 + 21)

$$\hat{\theta}_{yi} = \frac{N_{yi} + \alpha}{N_y + \alpha n}$$

Now we just multiply all the probabilities, and see who is bigger:

P(overall | positive) * P(liked | positive) * P(the | positive) * P(movie | positive) * P(positive) =
1.38 * 10^{-5} = 0.0000138
P(overall | negative) * P(liked | negative) * P(the | negative) * P(movie | negative) * P(negative) =
0.13 * 10^{-5} = 0.0000013

Our classifier gives “overall liked the movie” the positive tag.

Bernoulli Naive Bayes

- ✔ Bernoulli Naive Bayes is a subcategory of the Naive Bayes Algorithm. It is used for the classification of binary features such as 'Yes' or 'No', '1' or '0', 'True' or 'False' etc. The features are independent of one another.
- ✔ Bernoulli Naive Bayes is basically used for spam detection, text classification, Sentiment Analysis, used to determine whether a certain word is present in a document or not.

Bernoulli Naive Bayes Classifier is based on the following rule:

$$P(x_i \mid y) = P(i \mid y)x_i + (1 - P(i \mid y))(1 - x_i)$$

- ✔ Here, $p(x_i \mid y)$ is the conditional probability of x_i occurring provided y has occurred.
- ✔ i is the event
- ✔ x_i holds binary value either 0 or 1

Bernoulli Naive Bayes

✔ If X is random variable and is Bernoulli-distributed, it can assume only two values (for simplicity, let's call them 0 and 1) and their probability is:

$$P(X) = \begin{cases} p & if\ X = 1 \\ q & if\ X = 0 \end{cases}$$

$$where\ q = 1 - p\ and\ 0 < p < 1$$

Now, let us solve a problem for Bernoulli Naive Bayes:

Example 01

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

- ✔ We want to classify an **instance 'X' with Confident=Yes, Studied=Yes and Sick=No.** So, first we need to calculate the class probabilities i.e **P(Pass)=3/5 and P(Fail)=2/5**
- ✔ Calculate individual probability with respect to each features. For example,
  P(Confident=Yes| Result=Pass) = 2/3
  P(Studied=Yes| Result=Pass) = 2/3
  P(Sick=No| Result=Pass) = 1/3
- ✔ Similarly,
- ✔ P(Confident=Yes| Result=Fail) = 1/2
  P(Studied=Yes| Result=Fail) = 1/2
  P(Sick=No| Result=Fail) = 1/2

Bernoulli Naive Bayes

- ✔ Now we calculate,

  P(X|Result=Pass)xP(Result=Pass) = (2/3) * (2/3) * (1/3) * (3/5) = **0.088**

  P(X|Result=Fail)xP(Result=Fail) = (1/2) * (1/2) * (1/2) * (2/5) = **0.05**

- ✔ Next step is to calculate the probability of estimator

  P(X) = P(Confident=Yes)xP(Studied=Yes)xP(Sick=No) = (3/5) * (3/5) * (2/5) = **0.144**

- ✔ Finally,

  P(Result=Pass|X) = 0.088/0.144 = 0.611

  P(Result=Fail|X) = 0.05/0.144 = 0.34

- ✔ As 0.611>0.31, the instnace with 'Confident=Yes, Studied=Yes and Sick=No' has result as 'Pass'.

Bernoulli Naive Bayes

- ✔ Let us consider the example below to understand Bernoulli Naive Bayes:-

Example 02

Adult	Gender	Fever	Disease
Yes	Female	No	False
Yes	Female	Yes	True
No	Male	Yes	False
No	Male	No	True
Yes	Male	Yes	True

- ✔ In the above dataset, we are trying to … a disease or not based on their age, gender, and fever. Here, 'Diseas… e the features.
- ✔ All values are binary.

Bernoulli Naive Bayes

- ✔ We wish to classify an instance ‘X’ where Adult=’Yes’, Gender= ‘Male’, and Fever=’Yes’.
- ✔
- ✔ Firstly, we calculate the class probability, probability of disease or not.
- ✔
- ✔ P(Disease = True) = ⅗
- ✔
- ✔ P(Disease = False) = ⅖
- ✔
- ✔ Secondly, we calculate the individual probabilities for each feature.
- ✔
- ✔ P(Adult= Yes | Disease = True) = ⅔
- ✔ P(Gender= Male | Disease = True) = ⅔
- ✔ P(Fever= Yes | Disease = True) = ⅔

Adult	Gender	Fever	Disease
Yes	Female	No	False
Yes	Female	Yes	True
No	Male	Yes	False
No	Male	No	True
Yes	Male	Yes	True

Bernoulli Naive Bayes

Adult	Gender	Fever	Disease
Yes	Female	No	False
Yes	Female	Yes	True
No	Male	Yes	False
No	Male	No	True
Yes	Male	Yes	True

- ✔ P(Adult= Yes | Disease = False) = ½
- ✔ P(Gender= Male | Disease = False) = ½
- ✔ P(Fever = Yes | Disease = False) = ½
- ✔
- ✔ Now, we need to find out two probabilities:-
- ✔ (i) P(Disease= True | X) = (P(X | Disease= True) * P(Disease=True))/ P(X)
- ✔ (ii) P(Disease = False | X) = (P(X | Disease = False) * P(Disease= False))/P
- ✔
- ✔ P(Disease = True | X) = ((⅔ *⅔ * ⅔) * (⅗))/P(X) = (8/27 * ⅗) / P(X) = 0.17/P(X)
- ✔
- ✔ P(Disease = False | X) = [(½ * ½ * ½) * (⅖)] / P(X) = [⅛ * ⅖] / P(X) = 0.05/ P(X)

Bernoulli Naive Bayes

✔ Now, we calculate estimator probability:-
✔
✔ P(X) = P(Adult= Yes) * P(Gender = Male) * P(Fever = Yes)
✔ = ⅗ * ⅗ * ⅗ = 27/125 = 0.21
✔
✔ So we get finally:-
✔
✔ P(Disease = True | X) = 0.17 / P(X)
✔ = 0.17 / 0.21
✔ = 0.80 - (1)
✔
✔ P(Disease = False | X) = 0.05 / P(X)
✔ = 0.05 / 0.21
✔ = 0.23 - (2)
✔ Now, we notice that (1) > (2), the result of instance ‘X’ is ‘True’, i.e., the person has the disease.

Example 03

1. **Data Preparation:** Begin with a set of binary data. Each row signifies a data sample while columns represent features.

DATA PREPARATION

Feature 1	Feature 2	Class
1	0	0
0	1	0
1	1	0
0	0	0
1	0	1
0	1	1
0	0	1
1	1	1
1	0	0
0	1	0

- Each row in the table represents a data sample
- each column represents a binary feature.
- For instance, "Feature 1" represents the presence (1) or absence (0)

2. **Training Phase** — Class Priors: Compute the prior probabilities of each class based on your training data.

Calculate the prior probabilities for each class based on the training data:

- Total samples = 10
- Class 0 samples = 6
- Class 1 samples = 4

Thus:

$P\ (Class\ 0) = 6/10 = 0.6$

$P\ (Class\ 1) = 4/10 = 0.4$

TRAINING PHASE – CLASS PRIORS

Calculate the **prior probabilities** of each class (Class A and Class B) based on the training data

$$P(Class\ 0) = \frac{\text{Number of Class 0 samples}}{\text{Total number of samples}} = \frac{6}{10} = 0.6$$

$$P(Class\ 1) = \frac{\text{Number of Class 1 samples}}{\text{Total number of samples}} = \frac{4}{10} = 0.4$$

3. **Training Phase — Feature Priors:** Calculate conditional probabilities for each feature based on their presence (1) or absence (0).

Next, compute the conditional probabilities for each feature based on its presence (1) or absence (0):

For Feature 1:

1. With Class 0: 3 out of 6 samples have Feature 1 = 1

$$P(\mathit{Feature}\ 1{=}1|\mathit{Class}\ 0) = 3/6 = 0.5$$

2. With Class 1: 3 out of 6 samples have Feature 1 = 1

$$P\ (\mathit{Feature}\ 1{=}1|\mathit{Class}\ 1) = 2/4 = 0.5$$

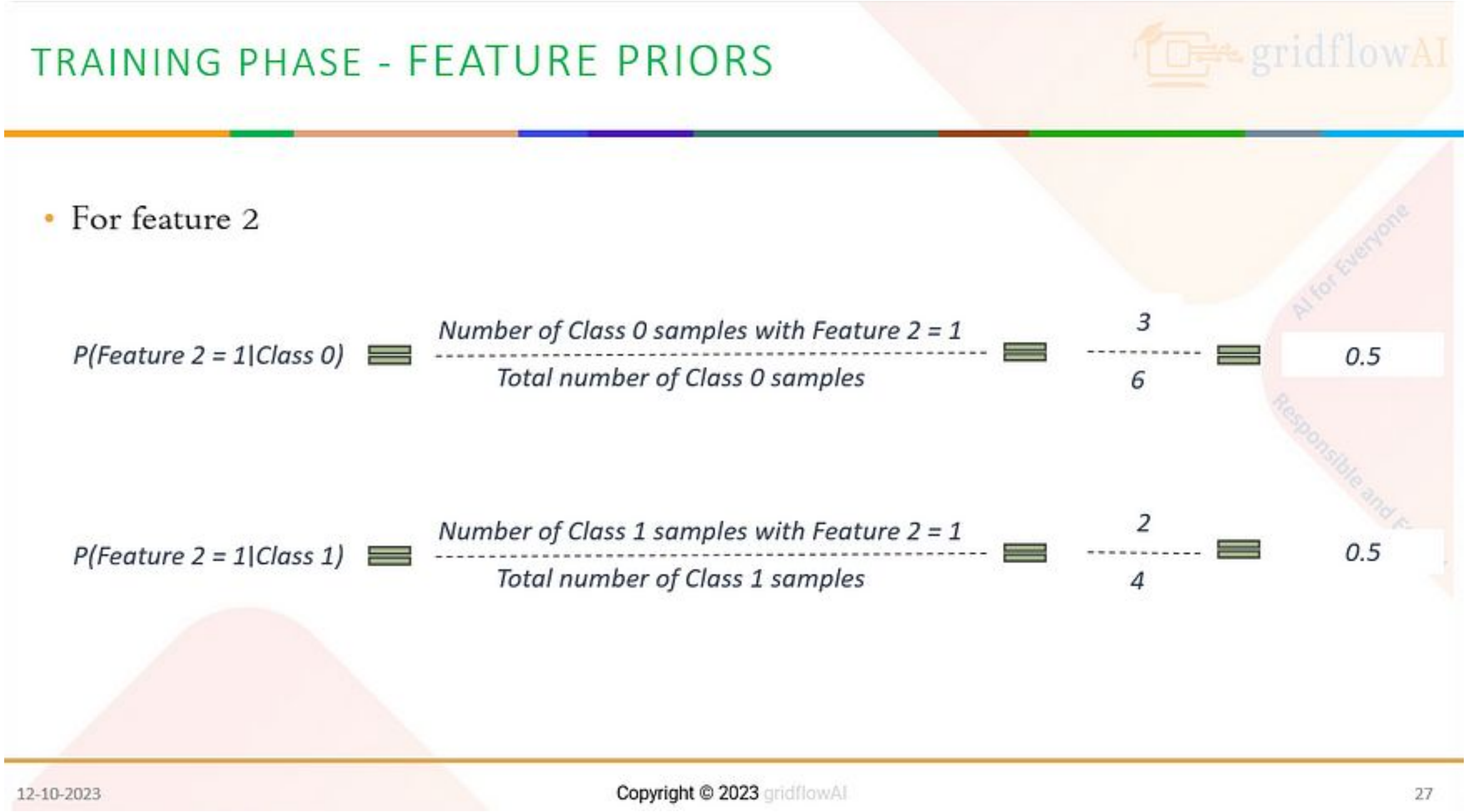
TRAINING PHASE - FEATURE PRIORS

- For each feature (Feature 1 and Feature 2), calculate conditional probabilities for each class based on the presence (1) or absence (0) of the feature
- For feature 1

$$P(\text{Feature } 1 = 1 | \text{Class } 0) = \frac{\text{Number of Class 0 samples with Feature } 1 = 1}{\text{Total number of Class 0 samples}} = \frac{3}{6}$$

$$P(\text{Feature } 1 = 1 | \text{Class } 1) = \frac{\text{Number of Class 1 samples with Feature } 1 = 1}{\text{Total number of Class 1 samples}} = \frac{2}{4}$$

- ✔ **For Feature 2:**
- ✔ With Class 0: 3 out of 6 samples have Feature 2= 1
- ✔ ***P*(*Feature* 2=1 | *Class* 0) = 3 /6 =0.5**
- ✔ 2. With Class 1: 2out of 4samples have Feature 2= 1
- ✔ ***P* (*Feature* 2=1 | *Class* 1)= 2/4 =0.5**



- ✔ 4. **Prediction:** Using the trained model, classify new samples.
- ✔ Calculate the likelihood of each class for the given features and compute the unnormalized posterior probability.
- ✔ **For Class 0:**
  P(Feature 1=1, Feature 2=0 | Class 0) = P(Feature 1=1 | Class 0) × P(Feature2=0 | Class 0)
  = 0.5 × 0.5
  = 0.25
- ✔ **Unnormalized Posterior**: P(Class 0) × Likelihood = 0.6 × 0.25 = 0.15

CALCULATE POSTERIORS

- Class 0

$$P(Class|Features) = \frac{P(Features|Class)*P(Class)}{P(Features)}$$

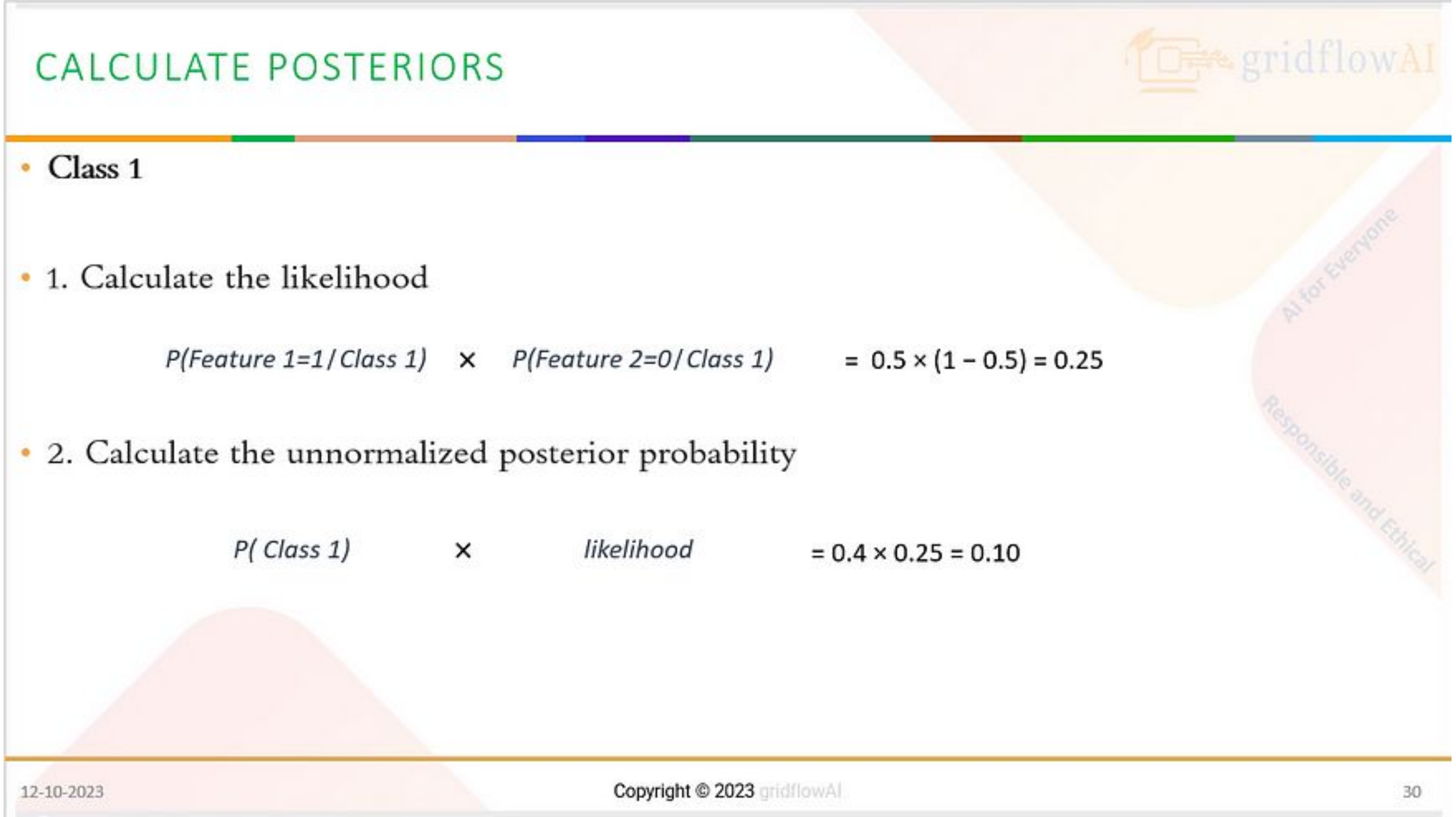
- 1. Calculate the likelihood

P(Feature 1=1 / Class 0) × *P(Feature 2=0 / Class 0)* = 0.5 × (1 – 0.5) = 0.25

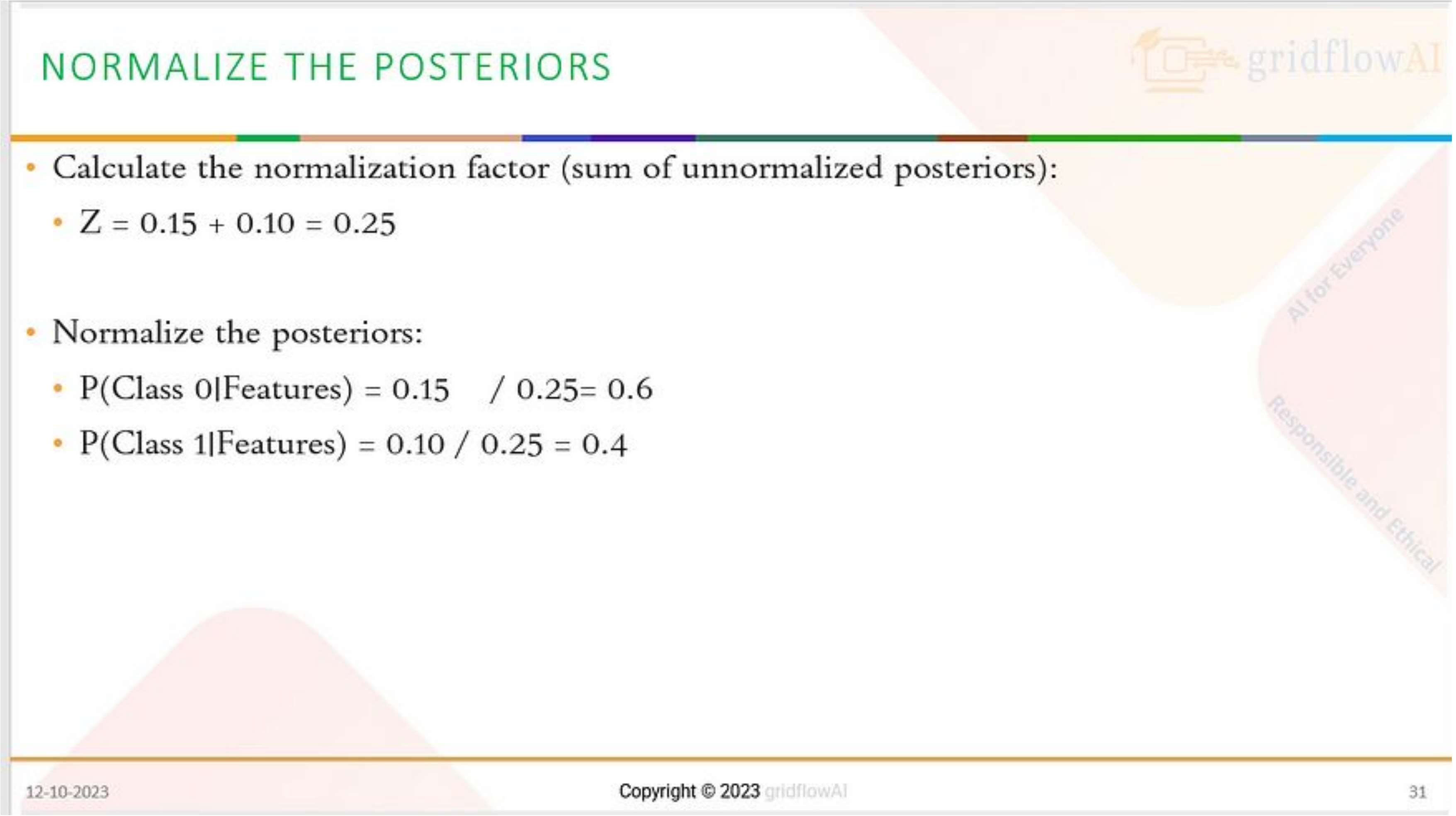
- 2. Calculate the unnormalized posterior probability

P(Class 0) × *likelihood* = 0.6 × 0.25 = 0.15

- **For Class 1:**
  P (Feature 1=1, Feature 2=0 | Class 1) = P (Feature 1=1 | Class 1) × P (Feature 2=0 | Class 1)
  = 0.5 × 0.5
  = 0.25
- **Unnormalized Posterior**: P (Class 1) × Likelihood = 0.4 × 0.25 = 0.10



- ✔ **Normalize these probabilities**
- ✔ Z = 0.15 (Class 0) + 0.10 (Class 1) = 0.25
- ✔ P (Class 0|Features) = 0.15 / 0.25 = 0.6
- ✔ P (Class 1|Features) = 0.10 / 0.25 = 0.4



DECIDE

- Since
  - $P(\text{Class } 0|\text{Features}) = 0.6 > P(\text{Class } 1|\text{Features}) = 0.4$
    - the <u>prediction</u> is Class 0.

Thank You

Example

Using the data, we have to identify the species of an entity with the following attributes.
X={Color=Green, Legs=2, Height=Tall, Smelly=No}
To predict the class label for the above attribute set, we will first calculate the probability of the species being M or H in total.
P(Species=M)=4/8=0.5 P(Species=H)=4/8=0.5
Next, we will calculate the conditional probability of each attribute value for each class label.
P(Color=White/Species=M)=2/4=0.5
P(Color=White/Species=H)=¾=0.75
P(Color=Green/Species=M)=2/4=0.5
P(Color=Green/Species=H)=¼=0.25 P(Legs=2/Species=M)=1/4=0.25
P(Legs=2/Species=H)=4/4=1 P(Legs=3/Species=M)=3/4=0.75
P(Legs=3/Species=H)=0/4=0 P(Height=Tall/Species=M)=3/4=0.75
P(Height=Tall/Species=H)=2/4=0.5 P(Height=Short/Species=M)=1/4=0.25
P(Height=Short/Species=H)=2/4=0.5 P(Smelly=Yes/Species=M)=3/4=0.75
P(Smelly=Yes/Species=H)=1/4=0.25 P(Smelly=No/Species=M)=1/4=0.25
P(Smelly=No/Species=H)=3/4=0.75

Sl. No.	Color	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

Example

- ✔ We can tabulate the above calculations in the tables for better visualization.
- ✔ The conditional probability table for the Color attribute is as follows.

Color	M	H
White	0.5	0.75
Green	0.5	0.25

- ✔ The conditional probability table for the Legs attribute is as follows.

Legs	M	H
2	0.25	1
3	0.75	0

Example

- ✔ The conditional probability table for the Height attribute is as follows.

Height	M	H
Tall	0.75	0.5
Short	0.25	0.5

- ✔ The conditional probability table for the Smelly attribute is as follows.

Smelly	M	H
Yes	0.75	0.25
No	0.25	0.75

Example

Now that we have calculated the conditional probabilities, we will use them to calculate the probability of the new attribute set belonging to a single class.
Let us consider X= {Color=Green, Legs=2, Height=Tall, Smelly=No}.
Then, the probability of X belonging to Species M will be as follows.
P(M/X)=P(Species=M)*P(Color=Green/Species=M)*P(Legs=2/Species=M)*P(Height=Tall/Species=M)*P(Smelly=No/Species=M) =0.5*0.5*0.25*0.75*0.25 =0.0117
Similarly, the probability of X belonging to Species H will be calculated as follows.
P(H/X)=P(Species=H)*P(Color=Green/Species=H)*P(Legs=2/Species=H)*P(Height=Tall/Species=H)*P(Smelly=No/Species=H) =0.5*0.25*1*0.5*0.75 =0.0468
So, the probability of X belonging to Species M is 0.0117 and that to Species H is 0.0468.
Hence, we will assign the entity X with attributes {Color=Green, Legs=2, Height=Tall, Smelly=No} to species H.