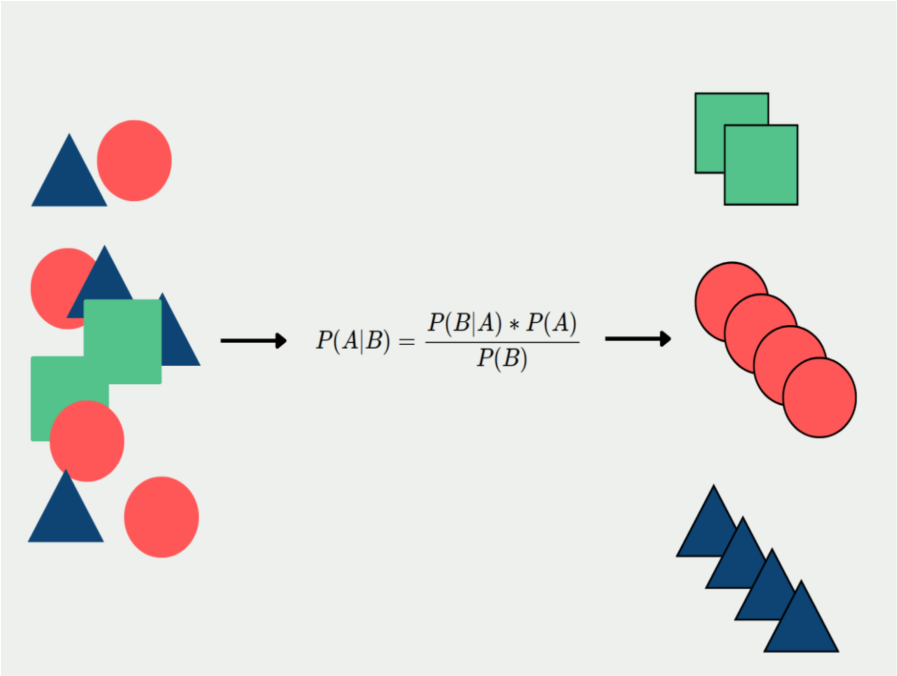
COMP30027 MACHINE LEARNING TUTORIAL

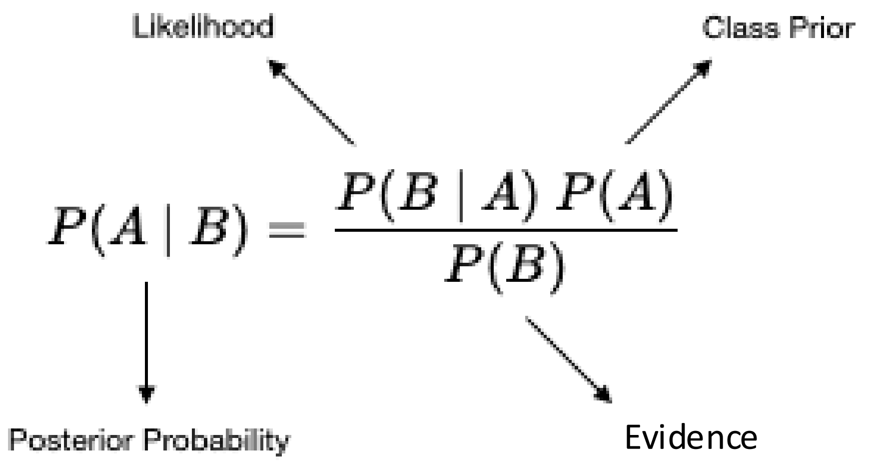
Workshop - 3

Naive Bayes



Naive Bayes is a simple yet powerful probabilistic classifier. It's known for its efficiency and effectiveness in various machine learning applications. These include spam filtering, text classification, etc. In some cases, it outperforms complex models, especially on smaller datasets.

Understanding the Naive Bayes Equation



Bayes Theorem

The Naive Bayes equation is P(A|B) = [P(B|A) * P(A)] / P(B). This calculates the probability of hypothesis A given evidence B.

Components

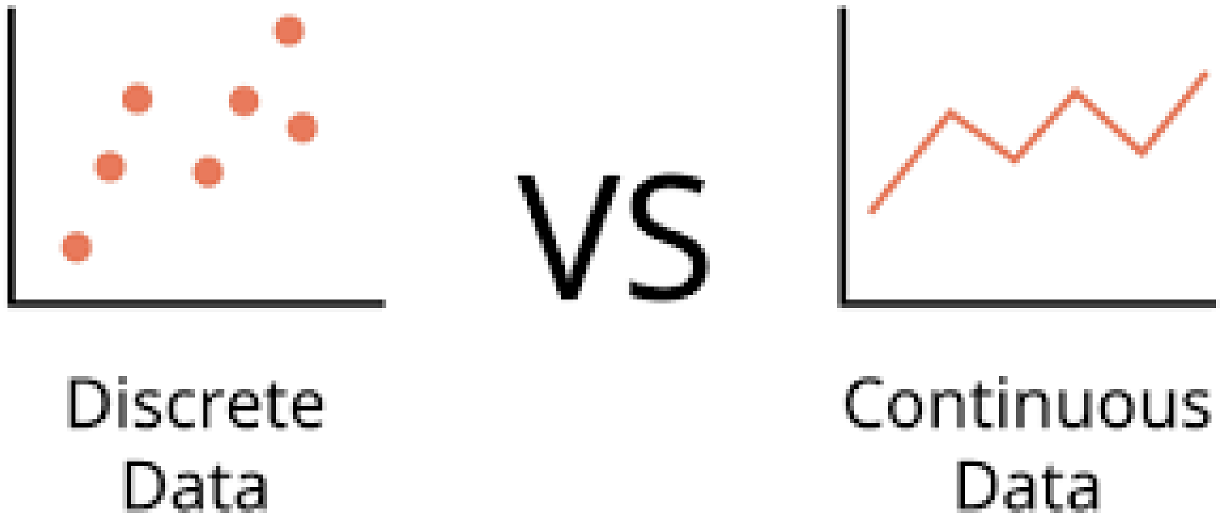
- Posterior: Probability of A given B
- Prior: Probability of A before evidence
- Likelihood: Probability of B given A
- Evidence: Probability of B

Key Assumption

Feature independence is assumed, which is why it's "Naive".

Discrete vs. Continuous

Data comes in different forms. Naive Bayes handles them differently. For continuous data, discretization or Gaussian Naive Bayes is used.



Discrete

It consists of distinct, separate values and does not take fractional or decimal values between them.

Continuous

A **continuous value** can take an infinite number of values within a given range.

Q1

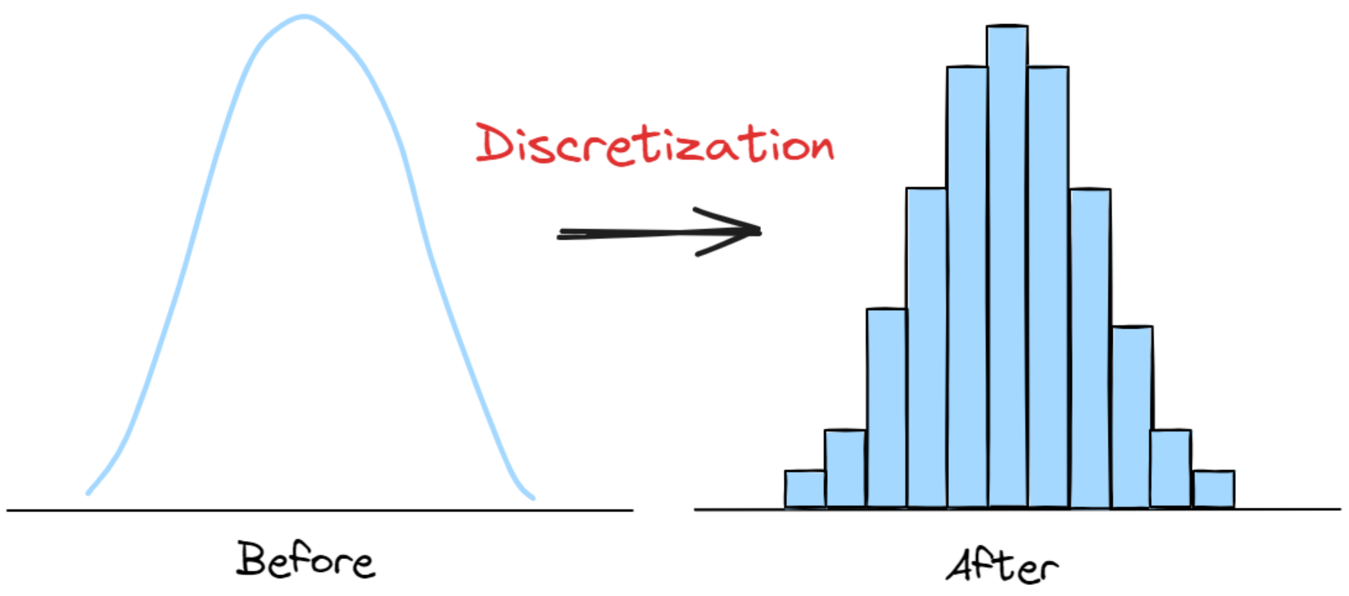
Which attribute is discrete and which is continuous?

Discrete vs. Continuous

ID	Color	Weight (g)
	Discrete	Continuous
1	Red	1021.2
2	Red	1027.0
3	Red	1012.5
4	Blue	1010.4
5	Blue	1019.5
6	Gold	1016.4
7	Gold	995.4
8	Gold	1012.8

Discretisation

Discretisation is the process of converting a **continuous variable** (temperature, weight) into a **discrete one** by dividing its range into **finite intervals** ("low", "medium", and "high").



Three discretisation techniques are:

Equal Width

Divides the range of values into bins of equal size (equal width)

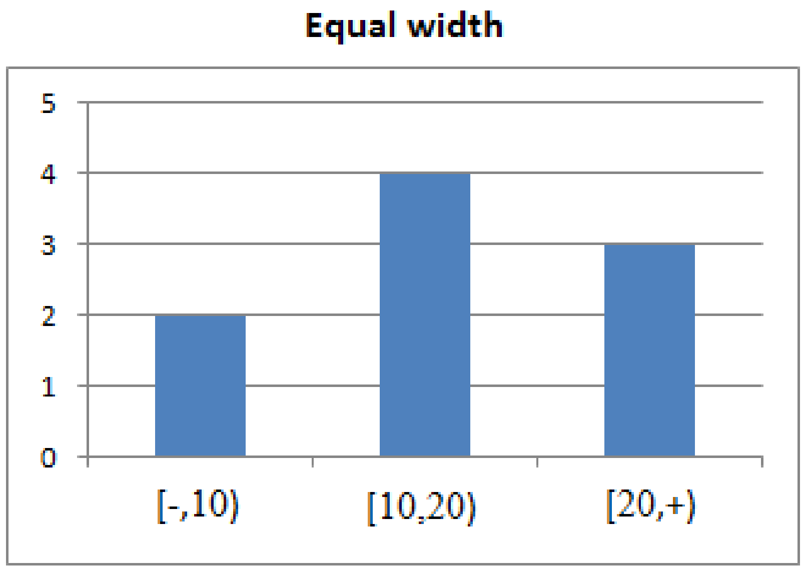
Equal Frequency

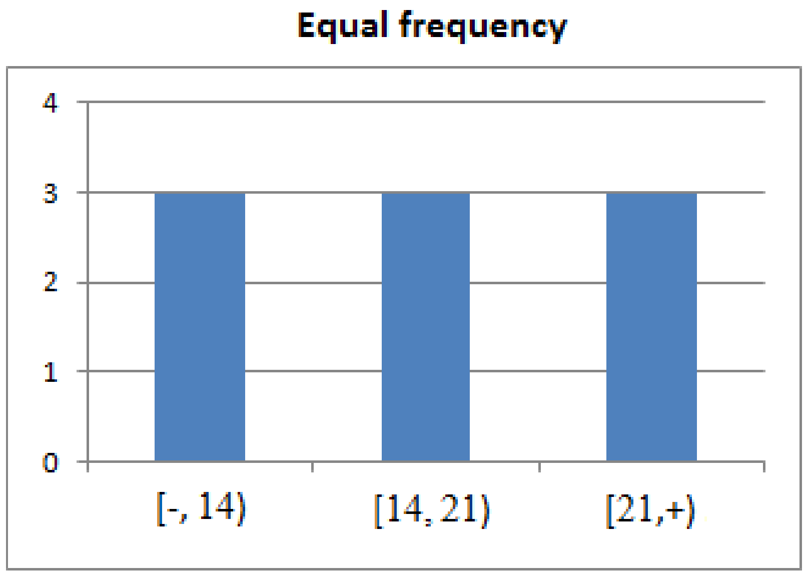
Each bin contains approximately the same number of data points.

K-means

Convert continuous values into groups (e.g., Salary into low, medium, high)

Discretisation





Equal Width

Divides the range of values into bins of equal size (equal width)

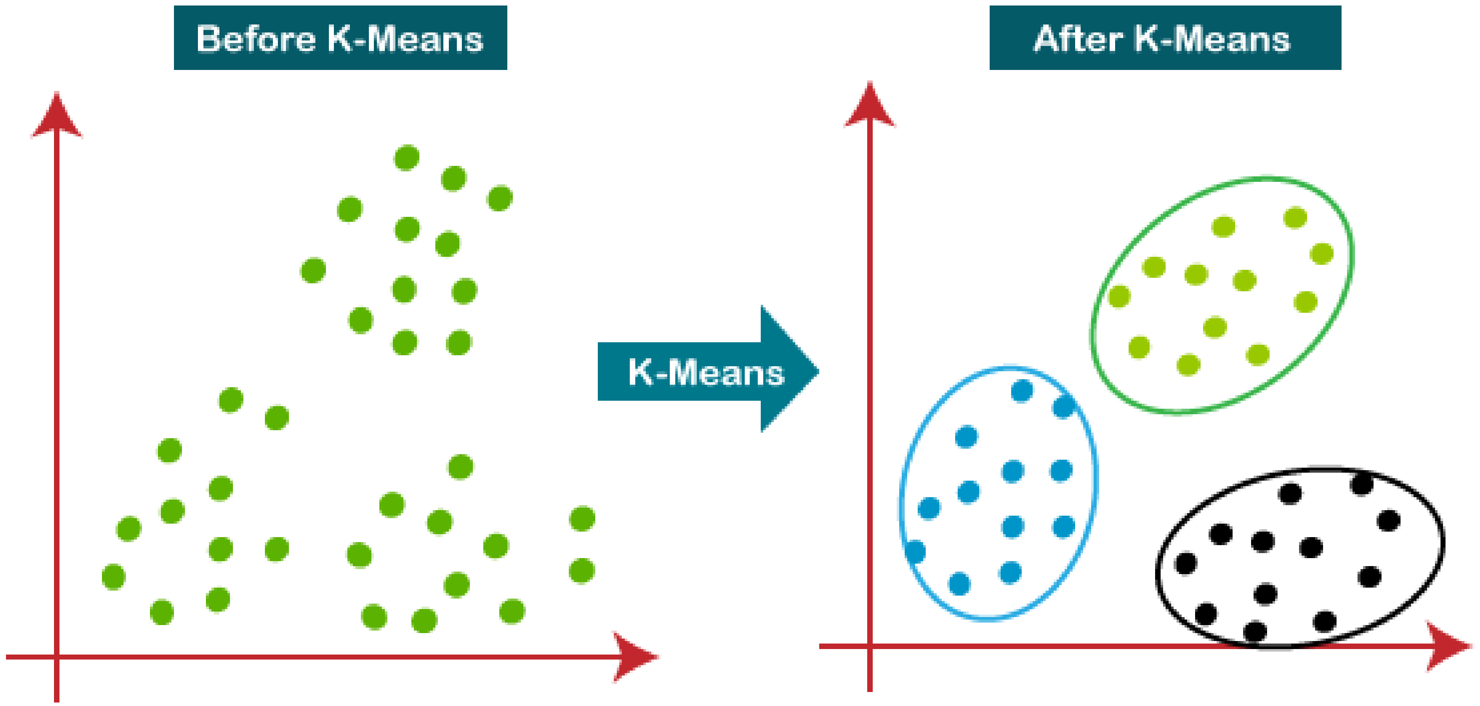
Equal Frequency

Each bin contains approximately the same number of data points.

K-means

Convert continuous values into groups (e.g., Salary into low, medium, high)

Discretisation



Equal Width

Divides the range of values into bins of equal size (equal width

Equal Frequency

Each bin contains approximately the same number of data points.

K-means

Convert continuous values into groups (e.g., Salary into low, medium, high)

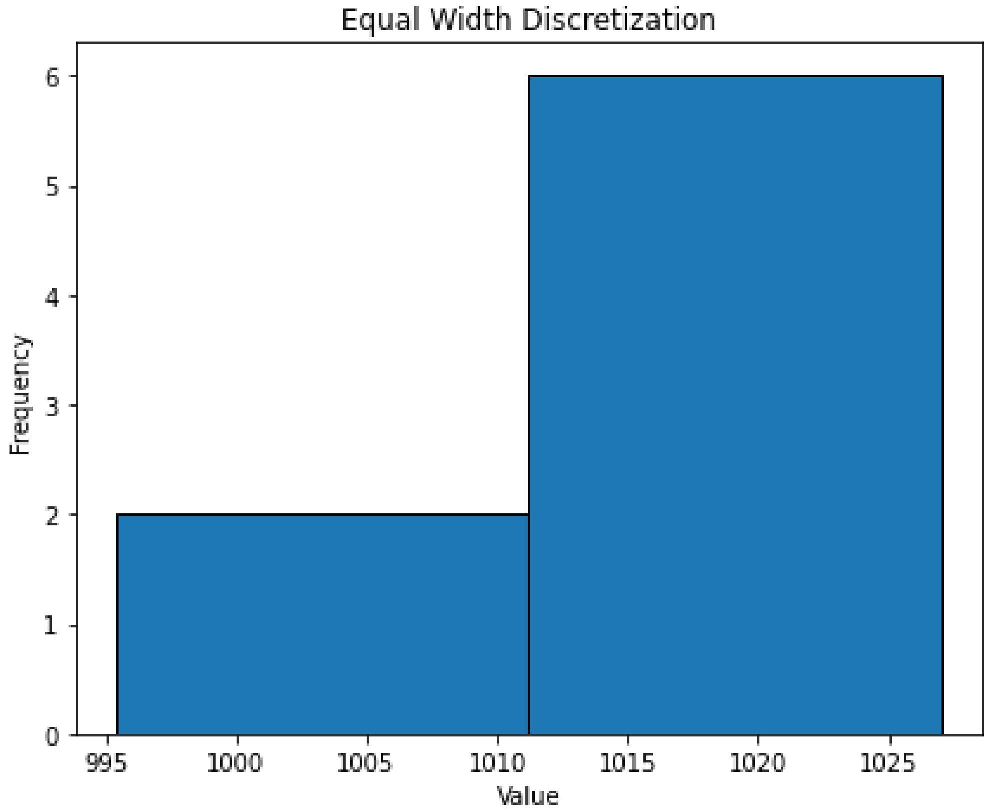
Discretisation

Q2

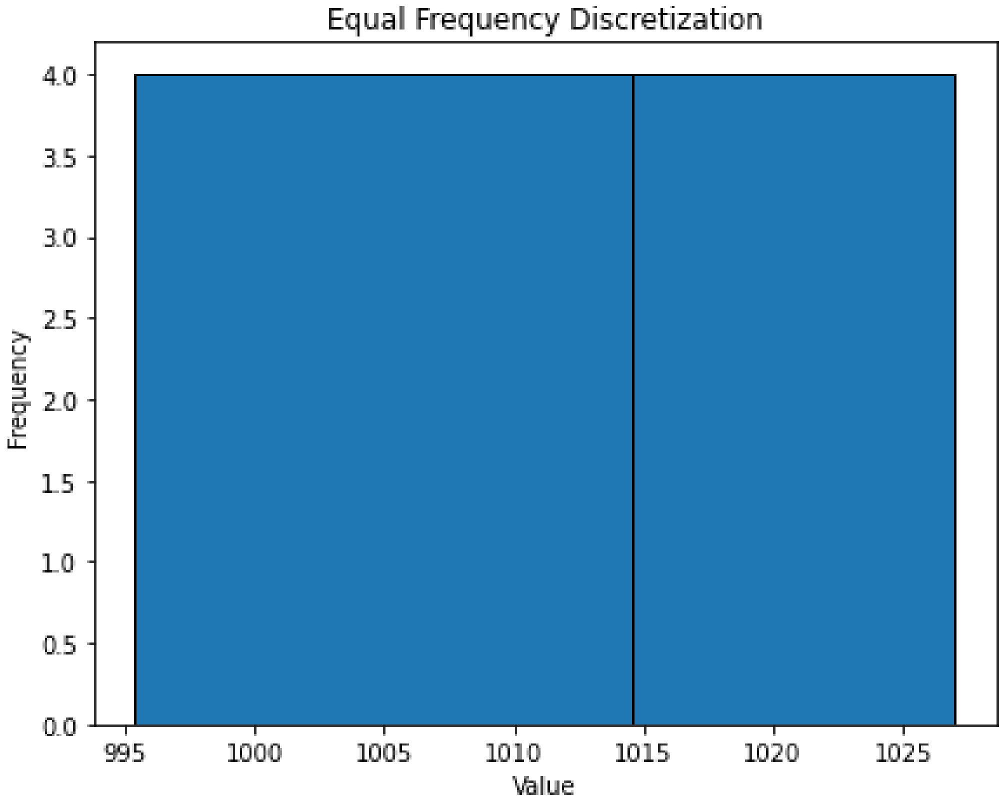
Discretise the continuous attribute into 2 bins using the (unsupervised) methods of **equal width**, **equal frequency**, and **k-means** (break ties where necessary).

	Discrete	Continuous
ID	**Color**	**Weight (g)**
1	Red	1021.2
2	Red	1027.0
3	Red	1012.5
4	Blue	1010.4
5	Blue	1019.5
6	Gold	1016.4
7	Gold	995.4
8	Gold	1012.8

Equal Width



Equal Frequency



ID	Color	Weight (g)
1	Red	1021.2
2	Red	1027.0
3	Red	1012.5
4	Blue	1010.4
5	Blue	1019.5
6	Gold	1016.4
7	Gold	995.4
8	Gold	1012.8

Step-by-Step Discretisation Using K-Means (k = 2)

- Initialize Two Random Centroids
- Assign Each Point to the Nearest Centroid
- Compute New Centroids
- Reassign Each Point to the New Centroids

ID	Color	Weight (g)
1	Red	1021.2
2	Red	1027.0
3	Red	1012.5
4	Blue	1010.4
5	Blue	1019.5
6	Gold	1016.4
7	Gold	995.4
8	Gold	1012.8

- **Initialize Two Random Centroids**

We randomly choose two initial centroids. Let's assume:

- **Centroid 1:** 1021.2
- **Centroid 2:** 995.4

• Assign Each Point to the Nearest Centroid

Weight	Distance to Centroid 1 (1021.2)	Distance to Centroid 2 (995.4)	Assigned Cluster
1021.2	0	25.8	1
1027.0	5.8	31.6	1
1012.5	8.7	17.1	1
1010.4	10.8	15.0	1
1019.5	1.7	24.1	1
1016.4	4.8	21.0	1
995.4	25.8	0	2
1012.8	8.4	17.4	1

• Compute New Centroids

New centroid for each cluster is the **mean** of the assigned weights.

•**New Centroid 1:** (1021.2+1027.0+1012.5+1010.4+1019.5+1016.4+1012.8)/7=

1017.11

•**New Centroid 2:** Since only **995.4** is in **Cluster 2**, the centroid remains **995.4**.

- **Reassign Each Point to the New Centroids**

Weight	Distance to New Centroid 1 (1017.11)	Distance to Centroid 2 (995.4)	Assigned Cluster
1021.2	4.09	25.8	1
1027.0	9.89	31.6	1
1012.5	4.61	17.1	1
1010.4	6.71	15.0	1
1019.5	2.39	24.1	1
1016.4	0.71	21.0	1
995.4	21.71	0	2
1012.8	4.31	17.4	1

No change from the previous assignment of clusters, so we can stop

Bin 1 (Centroid = 1017.11): {1021.2, 1027.0, 1012.5, 1010.4, 1019.5, 1016.4, 1012.8}

Bin 2 (Centroid = 995.4): {995.4}

Q3

How could the discrete variable be converted to a continuous numeric variable?

ID	R	B	G	Weight (g)
1				1021.2
2				1027.0
3				1012.5
4				1010.4
5				1019.5
6				1016.4
7				995.4
8				1012.8

ID	Color	Weight (g)
1	Red	1021.2
2	Red	1027.0
3	Red	1012.5
4	Blue	1010.4
5	Blue	1019.5
6	Gold	1016.4
7	Gold	995.4
8	Gold	1012.8

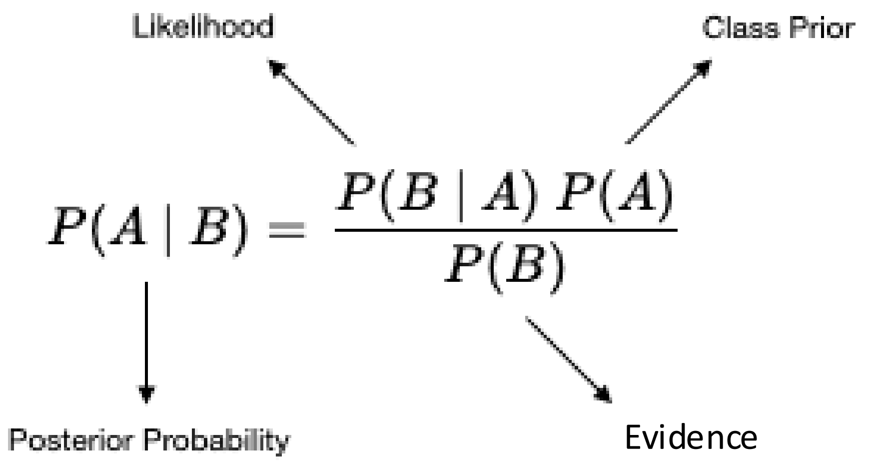
Q3

How could the discrete variable be converted to a continuous numeric variable?

ID	R	B	G	Weight (g)
1	1	0	0	1021.2
2	1	0	0	1027.0
3	1	0	0	1012.5
4	0	1	0	1010.4
5	0	1	0	1019.5
6	0	0	1	1016.4
7	0	0	1	995.4
8	0	0	1	1012.8

R – 1 0 0
B – 0 1 0
G – 0 0 1

Naive Bayes Equation



Bayes Theorem

The Naive Bayes equation is P(A|B) = [P(B|A) * P(A)] / P(B). This calculates the probability of hypothesis A given evidence B.

Components

- Posterior: Probability of A given B
- Prior: Probability of A before evidence
- Likelihood: Probability of B given A
- Evidence: Probability of B

Key Assumption

Feature independence is assumed, which is why it's "Naive".

Naive Bayes Classifier

T= < x_1 , $x_{2,}$ . . . . ., x_n > ← Instance

A ← Class

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

↓ Naïve assumption: Feature independence

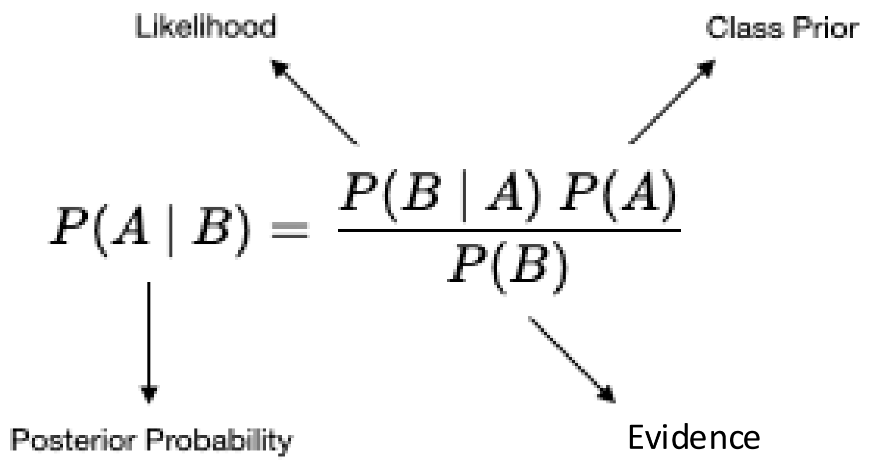
Naive Bayes ⟶ $$P(A \mid x_1, \ldots, x_n) = P(x_1 \mid A) \cdot P(x_2 \mid A) \cdot P(x_i \mid A)\, P(A)$$

$$\hat{A} = \underset{A}{argmax}\, P(A) \prod_{i=1}^{n} P(x_i \mid A)$$

Classifying Data Using Naive Bayes

Classifying data with Naive Bayes involves calculating prior and likelihood probabilities. Then, apply the Naive Bayes equation. Finally, assign the data point to the class with the highest posterior probability.

1. **Calculate priors**
2. **Calculate likelihoods**
3. **Compute posteriors**
4. **Assign class**



Bayes Theorem

Naive Bayes for Cold/Flu Diagnosis

- Given a training data set, what probabilities do we need to estimate?

Headache	Sore	Temperature	Cough	Diagnosis
severe	mild	high	yes	**Flu**
no	severe	normal	yes	**Cold**
mild	mild	normal	yes	**Flu**
mild	no	normal	no	**Cold**
severe	severe	normal	yes	**Flu**

We need $P(c_j)$, $P(x_i|c_j)$ for every x_i, c_j

Naive Bayes for Cold/Flu Diagnosis

Headache	Sore	Temperature	Cough	Diagnosis
severe	mild	high	yes	**Flu**
no	severe	normal	yes	**Cold**
mild	mild	normal	yes	**Flu**
mild	no	normal	no	**Cold**
severe	severe	normal	yes	**Flu**

P(Flu) = 3/5
P(Headache = severe|Flu) = 2/3
P(Headache = mild|Flu) = 1/3
P(Headache = no|Flu) = 0/3

P(Cold) = 2/5
P(Headache = severe|Cold) = 0/2
P(Headache = mild|Cold) = 1/2
P(Headache = no|Cold) = 1/2

Naive Bayes for Cold/Flu Diagnosis

P(Flu) = 3/5	P(Cold) = 2/5
P(Headache = severe|Flu) = 2/3	P(Headache = severe|Cold) = 0/2
P(Headache = mild|Flu) = 1/3	P(Headache = mild|Cold) = 1/2
P(Headache = no|Flu) = 0/3	P(Headache = no|Cold) = 1/2
P(Sore = severe|Flu) = 1/3	P(Sore = severe|Cold) = 1/2
P(Sore = mild|Flu) = 2/3	P(Sore = mild|Cold) = 0/2
P(Sore = no|Flu) = 0/3	P(Sore = no|Cold) = 1/2
P(Temp = high|Flu) = 1/3	P(Temp = high|Cold) = 0/2
P(Temp = normal|Flu) = 2/3	P(Temp = normal|Cold) = 2/2
P(Cough = yes|Flu) = 3/3	P(Cough = yes|Cold) = 1/2
P(Cough = no|Flu) = 0/3	P(Cough = no|Cold) = 1/2

Naive Bayes for Cold/Flu Diagnosis

- A patient comes to the clinic with mild headache, severe soreness, normal temperature, and no cough. Are they more likely to have cold or flu?

Cold: $P(C)P(H = m|C)P(S = s|C)P(T = n|C)P(C = n|C)$

$$\frac{2}{5}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{2}{2}\right)\left(\frac{1}{2}\right) = 0.05$$

Flu: $P(F)P(H = m|F)P(S = s|F)P(T = n|F)P(C = n|F)$

$$\frac{3}{5}\left(\frac{1}{3}\right)\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)\left(\frac{0}{3}\right) = 0$$

Naive Bayes for Cold/Flu Diagnosis

- A patient comes to the clinic with severe headache, mild soreness, high temperature, and no cough. Are they more likely to have cold or flu?

Cold: $P(C)P(H = s|C)P(S = m|C)P(T = h|C)P(C = n|C)$

$$\frac{2}{5}\left(\frac{0}{2}\right)\left(\frac{0}{2}\right)\left(\frac{0}{2}\right)\left(\frac{1}{2}\right) = 0$$

Flu: $P(F)P(H = s|F)P(S = m|F)P(T = h|F)P(C = n|F)$

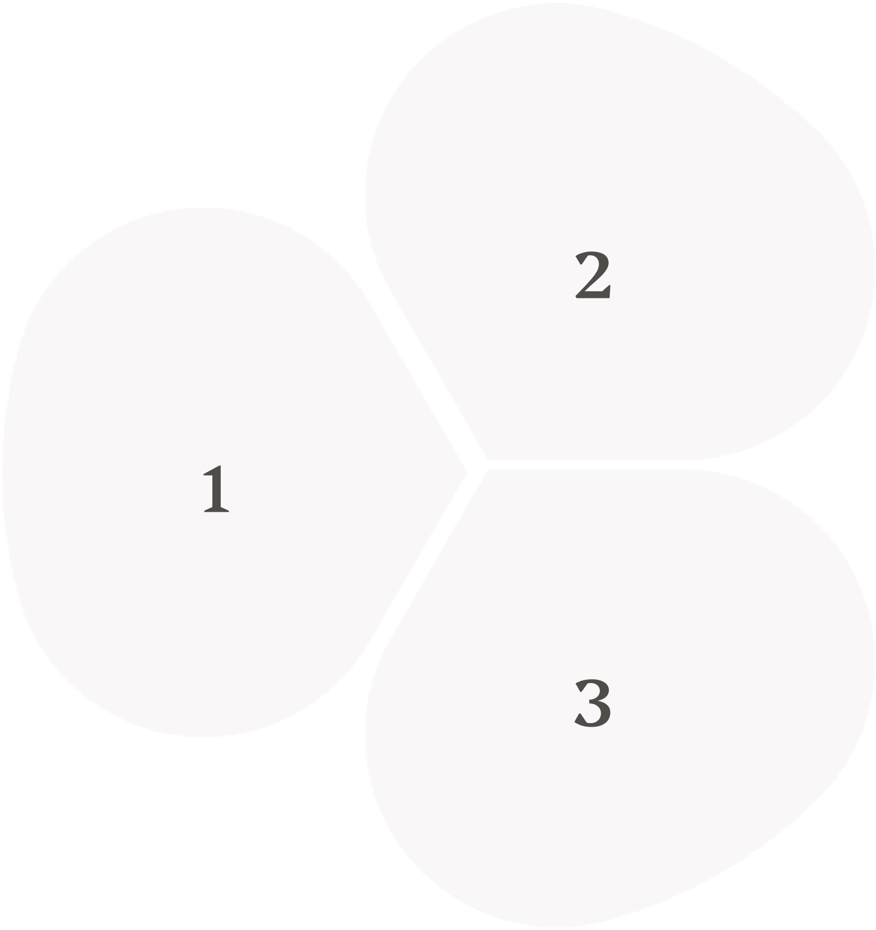
$$\frac{3}{5}\left(\frac{2}{3}\right)\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)\left(\frac{0}{3}\right) = 0$$

Probabilistic Smoothing

Smoothing is crucial in Naive Bayes to avoid the **zero-probability problem**. This problem occurs when a feature value doesn't appear in a class. For instance, in the training set of the previous example, the class ‘Cold’ doesn’t have any instances with the feature value “Headache = Severe.” So, the likelihood of observing a severe headache in a person with a Cold is zero. Smoothing techniques, like Laplace and Epsilon smoothing, help to solve this issue.

Zero Probability

Problem: feature absent in a class.



Laplace Smoothing

Adds 1 or α to each feature count.

Unsmoothed:

$$P_i = \frac{x_i}{N}$$

Smoothed:

$$P_i = \frac{x_i + \alpha}{N + \alpha d}$$

Epsilon Smoothing

Replace ‘0’ with a value ε $(0 < \varepsilon < 1)$.

Naïve Bayes

Given the following dataset, build a Naive Bayes model to predict the label "Play."

ID	Outlook	Temp	Humid	Wind	Play
A	S	H	N	F	N
B	S	H	H	T	N
C	O	H	H	F	Y
D	R	M	H	F	Y
E	R	C	N	F	Y
F	R	C	N	T	N

Q4

Classify the test instances using:

1. no smoothing
2. epsilon smoothing
3. Laplace smoothing (α=1)

Use the model to classify these test instances (? represents missing value):

ID	Outlook	Temp	Humid	Wind	Play
G	O	M	N	T	?
H	?	H	?	F	?

Solution: Naïve Bayes Model for Classification

We will follow these steps to classify the test instances using Naïve Bayes:

1.Compute prior probabilities

2.Compute likelihoods (conditional probabilities) for each feature

3.Classify test instances using:

1. No smoothing
2. Epsilon smoothing (ϵ = 0.001)
3. Laplace smoothing (α = 1)

Step 1: Compute Prior Probabilities

Prior probability of "Play = Y" and "Play = N":

$$P(Play = Y) = \frac{\text{count of Y}}{\text{total instances}} = \quad 3/6 = 0.5$$

$$P(Play = N) = \frac{\text{count of N}}{\text{total instances}} = \quad 3/6 = 0.5$$

Step 2: Compute Likelihoods (Conditional Probabilities)

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

Step 2: Compute Likelihoods (Conditional Probabilities)

$P(\text{Outlook} = S \mid \text{Play} = Y) =$ $\qquad$ $P(\text{Outlook} = S \mid \text{Play} = N) =$

$P(\text{Outlook} = O \mid \text{Play} = Y) =$ $\qquad$ $P(\text{Outlook} = O \mid \text{Play} = N) =$

$P(\text{Outlook} = R \mid \text{Play} = Y) =$ $\qquad$ $P(\text{Outlook} = R \mid \text{Play} = N) =$

$P(\text{Temp} = H \mid \text{Play} = Y) =$ $\qquad$ $P(\text{Temp} = H \mid \text{Play} = N) =$

$P(\text{Temp} = M \mid \text{Play} = Y) =$ $\qquad$ $P(\text{Temp} = M \mid \text{Play} = N) =$

$P(\text{Temp} = C \mid \text{Play} = Y) =$ $\qquad$ $P(\text{Temp} = C \mid \text{Play} = N) =$

$P(\text{Humid} = N \mid \text{Play} = Y) =$ $\qquad$ $P(\text{Humid} = N \mid \text{Play} = N) =$

$P(\text{Humid} = H \mid \text{Play} = Y) =$ $\qquad$ $P(\text{Humid} = H \mid \text{Play} = N) =$

$P(\text{Wind} = T \mid \text{Play} = Y) =$ $\qquad$ $P(\text{Wind} = T \mid \text{Play} = N) =$

$P(\text{Wind} = F \mid \text{Play} = Y) =$ $\qquad$ $P(\text{Wind} = F \mid \text{Play} = N) =$

Step 2: Compute Likelihoods (Conditional Probabilities)

ID	Outlook	Temp	Humid	Wind	Play
A	S	H	N	F	N
B	S	H	H	T	N
C	O	H	H	F	Y
D	R	M	H	F	Y
E	R	C	N	F	Y
F	R	C	N	T	N

$$P(\text{Outlook} = S \mid \text{Play} = Y) = \frac{\text{Count of instances with Outlook} = S \text{ and Play} = Y}{\text{Total count of instances with Play} = Y}$$

$$= \frac{0}{3}$$

$$P(\text{Outlook} = S \mid \text{Play} = N) = \frac{\text{Count of instances with Outlook} = S \text{ and Play} = N}{\text{Total count of instances with Play} = N}$$

$$= \frac{2}{3}$$

Step 2: Compute Likelihoods (Conditional Probabilities)

P(Outlook = S | Play = Y) = 0	P(Outlook = S | Play = N) = 2/3
P(Outlook = O | Play = Y) = 1/3	P(Outlook = O | Play = N) =
P(Outlook = R | Play = Y) =2/3	P(Outlook = R | Play = N) = 1/3
P(Temp = H | Play = Y) = 1/3	P(Temp = H | Play = N) = 2/3
P(Temp = M | Play = Y) = 1/3	P(Temp = M | Play = N) = 0
P(Temp = C | Play = Y) = 1/3	P(Temp = C | Play = N) = 1/3
P(Humid = N | Play = Y) = 1/3	P(Humid = N | Play = N) =2/3
P(Humid = H | Play = Y) = 2/3	P(Humid = H | Play = N) = 1/3
P(Wind = T | Play = Y) = 0	P(Wind = T | Play = N) = 2/3
P(Wind = F | Play = Y) = 1	P(Wind = F | Play = N) = 1/3

Step 3: Classify Test Instances

No smoothing

For **instance G**:

N: P(N) P(Outlook = 0|N) P(Temp = M|N) P(Humid = N|N) P(Wind = T|N) =

$$\frac{1}{2} \times 0 \times 0 \times \frac{2}{3} \times \frac{2}{3} = 0$$

Y: P(Y) P(Outlook = 0|Y) P(Temp = M|Y) P(Humid = N|Y) P(Wind = T|Y) =

$$\frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times 0 = 0$$

Tie. So, no label assigned

For **instance H**

N: P(N) P(Temp = H|N) P(Wind = F|N) =

$$\frac{1}{2} \times \frac{2}{3} \times \frac{1}{3} = \frac{1}{9}$$

Y: P(Y) P(Temp = H|Y) P(Wind = F|Y) =

$$\frac{1}{2} \times \frac{1}{3} \times 1 = \frac{1}{6}$$

✓

P(Outlook = S | Play = Y) = 0	P(Outlook = S | Play = N) = 2/3
P(Outlook = O | Play = Y) = 1/3	P(Outlook = O | Play = N) =
P(Outlook = R | Play = Y) =2/3	P(Outlook = R | Play = N) = 1/3
P(Temp = H | Play = Y) = 1/3	P(Temp = H | Play = N) = 2/3
P(Temp = M | Play = Y) = 1/3	P(Temp = M | Play = N) = 0
P(Temp = C | Play = Y) = 1/3	P(Temp = C | Play = N) = 1/3
P(Humid = N | Play = Y) = 1/3	P(Humid = N | Play = N) =2/3
P(Humid = H | Play = Y) = 2/3	P(Humid = H | Play = N) = 1/3
P(Wind = T | Play = Y) = 0	P(Wind = T | Play = N) = 2/3
P(Wind = F | Play = Y) = 1	P(Wind = F | Play = N) = 1/3

Step 3: Classify Test Instances

Epsilon smoothing (ϵ = 0.001)

For **instance G**:

N: P(N) P(Outlook = 0|N) P(Temp = M|N) P(Humid = N|N) P(Wind = T|N) =

$$\frac{1}{2} \times \epsilon \times \epsilon \times \frac{2}{3} \times \frac{2}{3} = \frac{2\epsilon^2}{9}$$

Y: P(Y) P(Outlook = 0|Y) P(Temp = M|Y) P(Humid = N|Y) P(Wind = T|Y) =

$$\frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \epsilon = \frac{\epsilon}{54}$$ ✓

For **instance H**

N: P(N) P(Temp = H|N) P(Wind = F|N) = $\frac{1}{2} \times \frac{2}{3} \times \frac{1}{3} = \frac{1}{9}$

Y: P(Y) P(Temp = H|Y) P(Wind = F|Y) = $\frac{1}{2} \times \frac{1}{3} \times 1 = \frac{1}{6}$ ✓

P(Outlook = S | Play = Y) = 0	P(Outlook = S | Play = N) = 2/3
P(Outlook = O | Play = Y) = 1/3	P(Outlook = O | Play = N) =
P(Outlook = R | Play = Y) =2/3	P(Outlook = R | Play = N) = 1/3
P(Temp = H | Play = Y) = 1/3	P(Temp = H | Play = N) = 2/3
P(Temp = M | Play = Y) = 1/3	P(Temp = M | Play = N) = 0
P(Temp = C | Play = Y) = 1/3	P(Temp = C | Play = N) = 1/3
P(Humid = N | Play = Y) = 1/3	P(Humid = N | Play = N) =2/3
P(Humid = H | Play = Y) = 2/3	P(Humid = H | Play = N) = 1/3
P(Wind = T | Play = Y) = 0	P(Wind = T | Play = N) = 2/3
P(Wind = F | Play = Y) = 1	P(Wind = F | Play = N) = 1/3

Step 3: Classify Test Instances

Laplace smoothing (α=1)

$$P_i = \frac{x_i + \alpha}{N + \alpha d}$$

The conditional probabilities with Laplace smoothing are:

P(Outlook = S | Play = Y) =	P(Outlook = S | Play = N) =
P(Outlook = O | Play = Y) =	P(Outlook = O | Play = N) =
P(Outlook = R | Play = Y) =	P(Outlook = R | Play = N) =
--	--
P(Temp = H | Play = Y) =	P(Temp = H | Play = N) =
P(Temp = M | Play = Y) =	P(Temp = M | Play = N) =
P(Temp = C | Play = Y) =	P(Temp = C | Play = N) =
--	--
P(Humid = N | Play = Y) =	P(Humid = N | Play = N) =
P(Humid = H | Play = Y) =	P(Humid = H | Play = N) =
--	--
P(Wind = T | Play = Y) =	P(Wind = T | Play = N) =
P(Wind = F | Play = Y) =	P(Wind = F | Play = N) =

Step 3: Classify Test Instances

Laplace smoothing $(\alpha{=}1)$

$$P(\text{Outlook} = S \mid \text{Play} = Y) = \frac{(\text{Count of instances with Outlook} = S \text{ and Play} = Y) + 1}{(\text{Total count of instances with Play} = Y) + (1 \times 3)}$$

$$= \frac{1}{6}$$

$$P(\text{Outlook} = S \mid \text{Play} = N) = \frac{(\text{Count of instances with Outlook} = S \text{ and Play} = Y) + 1}{(\text{Total count of instances with Play} = Y) + (1 \times 3)}$$

$$= \frac{3}{6}$$

$(\alpha{=}1)$

$$P_i = \frac{x_i + \alpha}{N + \alpha d}$$

No. of levels or distinct values in that Feature

'd' value for 'Outlook' = 3
'd' value for 'Humid' = 2

ID	Outlook	Temp	Humid	Wind	Play
A	S	H	N	F	N
B	S	H	H	T	N
C	O	H	H	F	Y
D	R	M	H	F	Y
E	R	C	N	F	Y
F	R	C	N	T	N

Step 3: Classify Test Instances

Laplace smoothing (α=1)

The conditional probabilities with Laplace smoothing are:

P(Outlook = S | Play = Y) = 1/6	P(Outlook = S | Play = N) = 3/6
P(Outlook = O | Play = Y) =2/6	P(Outlook = O | Play = N) =1/6
P(Outlook = R | Play = Y) = 3/6	P(Outlook = R | Play = N) =2/6
--	--
P(Temp = H | Play = Y) = 2/6	P(Temp = H | Play = N) =3/6
P(Temp = M | Play = Y) =2/6	P(Temp = M | Play = N) =1/6
P(Temp = C | Play = Y) =2/6	P(Temp = C | Play = N) =2/6
--	--
P(Humid = N | Play = Y) =2/3	P(Humid = N | Play = N) =3/5
P(Humid = H | Play = Y) =3/5	P(Humid = H | Play = N) =2/5
--	--
P(Wind = T | Play = Y) =1/5	P(Wind = T | Play = N) =3/5
P(Wind = F | Play = Y) =4/5	P(Wind = F | Play = N) =2/5

For **instance G**:

N: P(N) P(Outlook = 0|N) P(Temp = M|N) P(Humid = N|N) P(Wind = T|N) =

$$\frac{1}{2} \times \frac{1}{6} \times \frac{1}{6} \times \frac{3}{5} \times \frac{3}{5} = 0.005$$ ✓

Y: P(Y) P(Outlook = 0|Y) P(Temp = M|Y) P(Humid = N|Y) P(Wind = T|Y) =

$$\frac{1}{2} \times \frac{2}{6} \times \frac{2}{6} \times \frac{2}{5} \times \frac{1}{5} = 0.0044$$

For **instance H**

N: P(N) P(Temp = H|N) P(Wind = F|N) = $\frac{1}{2} \times \frac{3}{6} \times \frac{2}{5} = 0.1$

Y: P(Y) P(Temp = H|Y) P(Wind = F|Y) = $\frac{1}{2} \times \frac{2}{6} \times \frac{4}{5} = 0.13$ ✓

P(Outlook = S | Play = Y) = 1/6	P(Outlook = S | Play = N) = 3/6
P(Outlook = O | Play = Y) =2/6	P(Outlook = O | Play = N) =1/6
P(Outlook = R | Play = Y) = 3/6	P(Outlook = R | Play = N) =2/6
P(Temp = H | Play = Y) = 2/6	P(Temp = H | Play = N) =3/6
P(Temp = M | Play = Y) =2/6	P(Temp = M | Play = N) =1/6
P(Temp = C | Play = Y) =2/6	P(Temp = C | Play = N) =2/6
P(Humid = N | Play = Y) =2/3	P(Humid = N | Play = N) =3/5
P(Humid = H | Play = Y) =3/5	P(Humid = H | Play = N) =2/5
P(Wind = T | Play = Y) =1/5	P(Wind = T | Play = N) =3/5
P(Wind = F | Play = Y) =4/5	P(Wind = F | Play = N) =2/5

Naïve Bayes: Strength and Weakness

Strength

- **Simple & Fast**: Easy to implement and computationally efficient even on large datasets.
- **Requires Less Training Data**: Works well with relatively small amounts of labelled data.
- **Performs Well in High Dimensions**: Effective when you have many features (e.g., text classification).
- **Works Well with Categorical Data**: Naturally handles discrete features like word counts or categories.
- **Can Handle Missing Data**: Gracefully deals with missing values using smoothing techniques.

Weakness

- **Strong Independence Assumption**: Assumes all features are conditionally independent, which is rarely true in practice.
- **Zero Probability Problem**: If a feature value is unseen during training, it leads to zero probability (solved with smoothing).
- **Can Be Overconfident**: Outputs extreme probabilities even when uncertain.