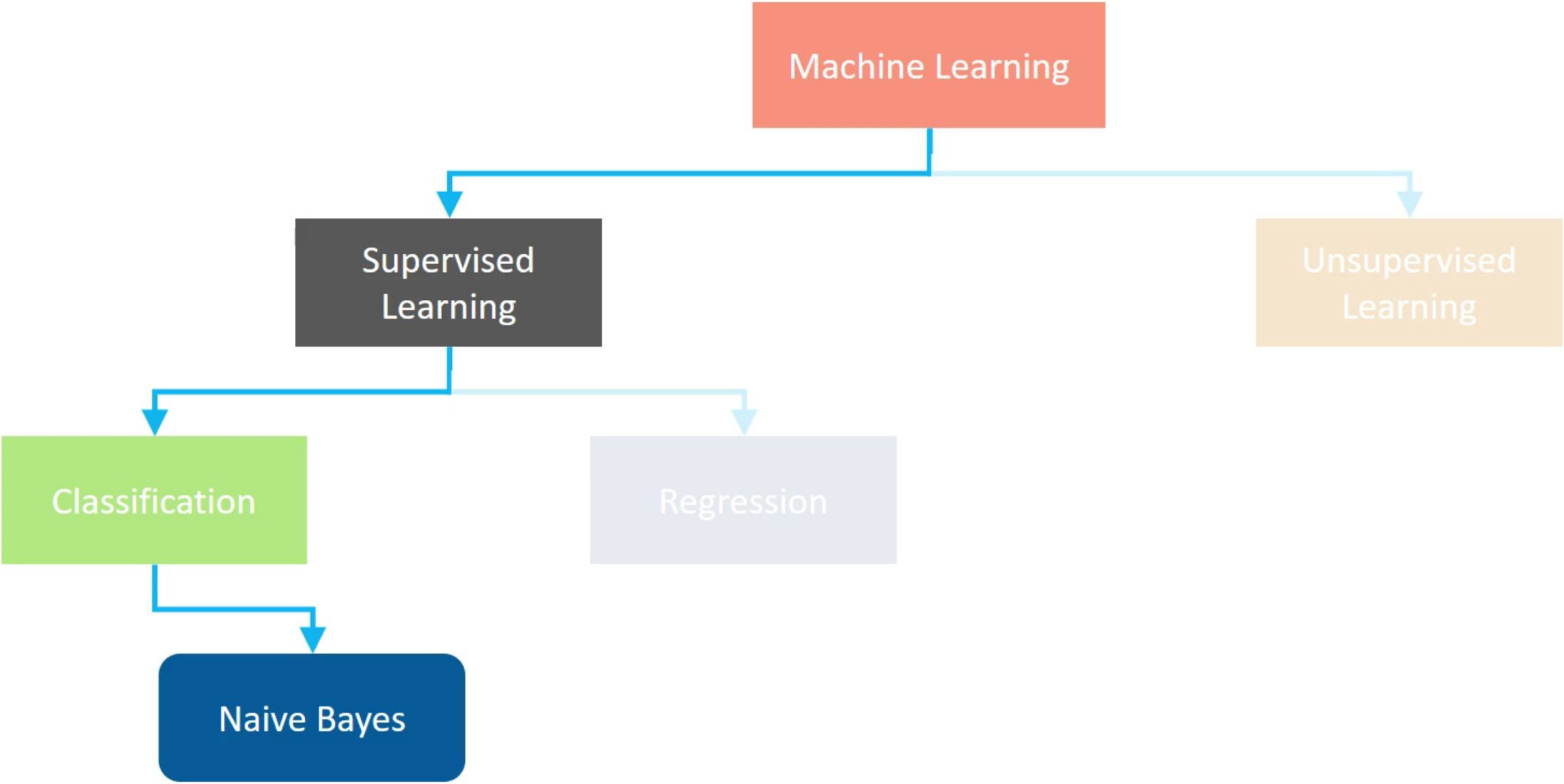
Naive Bayes - Classification Algo

Prepared By: Dr.Mydhili K Nair, Professor, ISE Dept, RIT
For: Machine Learning Elective Class
Target Audience: Sem 6 Students
Term: Feb to June 2019

- What is Naive Bayes?
- Naive Bayes and Machine Learning
- Why do we need Naive Bayes?
- Understanding Naive Bayes Classifier
- Advantages of Naive Bayes Classifier
- Demo – Text Classification using Naive Bayes

Python Code demo - not in the slides


Machine Learning
Supervised Learning
Unsupervised Learning
Classification
Regression
Naive Bayes

Classification
Frequency Table
ZeroR
One R
Naïve Bayesian
Decision Tree
Covariance Matrix
Linear Discriminant Analysis
Logistic Regression
Similarity Functions
K Nearest Neighbours
Others
Artificial Neural Network
Support Vector Machine

Derivation of Bayes' Theorem

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

Cumulative Rule

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

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

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

Bayes' Theorem for Naive Bayes Algorithm

In a machine learning classification problem, there are multiple features and classes, say, C_1, C_2, ……, C_k . The main aim in the Naive Bayes algorithm is to calculate the conditional probability of an object with a feature vector x_1, x_2,….., x_n belongs to a particular class C_i,

Bat = {Made of Wood, White or Beige, cylindrical hand holder }

Feature Vector for **Bat** & **Ball**

Ball = {round, can be hit, rolls }

Bayes' Theorem for Naive Bayes Algorithm

In a machine learning classification problem, there are multiple features and classes, say, C_1, C_2, ..., C_k . The main aim in the Naive Bayes algorithm is to calculate the conditional probability of an object with a feature vector x_1, x_2,....., x_n belongs to a particular class C_i,

Cricket Bat = {Made of Wood, White or Beige, Hit portion is Long rectangle, cylindrical hand holder }

Feature Vector for **Cricket Bat** & **Table Tennis Bat**

Table Tennis Bat = {Made of Wood, White or Beige, Hit portion is circular, cylindrical hand holder }

Bayes' Theorem for Naive Bayes Algorithm

In a machine learning classification problem, there are multiple features and classes, say, C_1, C_2, ……, C_k . The main aim in the Naive Bayes algorithm is to calculate the conditional probability of an object with a feature vector x_1, x_2,….., x_n belongs to a particular class C_i,

$$P(C_i|x_1, x_2, \ldots, x_n) = \frac{P(x_1, x_2, \ldots, x_n|C_i).P(C_i)}{P(x_1, x_2, \ldots, x_n)} \text{ for } 1 \leq i \leq k$$

Now, the numerator of the fraction on right-hand side of the equation above is

$$P(x_1, x_2, \ldots, x_n | C_i).P(C_i) = P(x_1, x_2, \ldots, x_n, C_i)$$

$$\begin{aligned} P(x_1, x_2, \ldots, x_n, C_i) &= P(x_1 | x_2, \ldots, x_n, C_i).P(x_2, \ldots, x_n, C_i) \\ &= P(x_1 | x_2, \ldots, x_n, C_i).P(x_2 | x_3, \ldots, x_n, C_i)P(x_3, \ldots, x_n, C_i) \\ &= \ldots \\ &= P(x_1 | x_2, \ldots, x_n, C_i).P(x_2 | x_3, \ldots, x_n, C_i) \ldots P(x_{n-1} | x_n, C_i).P(x_n | C_i).P(C_i) \end{aligned}$$

The conditional probability term, $P(x_j | x_{\{j+1\}}, \ldots, x_n, C_i)$ becomes $P(x_j | C_i)$ because of the assumption that features are independent

From the calculation above and the independence assumption, the Bayes theorem boils down to the following easy expression

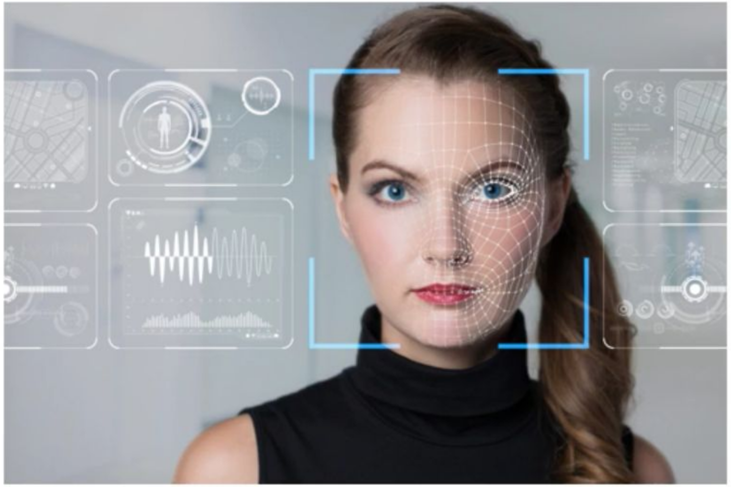
$$P(C_i|x_1, x_2, \ldots, x_n) = \left(\prod_{j=1}^{j=n} P(x_j|C_i)\right) \cdot \frac{P(C_i)}{P(x_1, x_2, \ldots, x_n)} \text{ for } 1 \leq i \leq k$$

The expression $P(x_1, x_2, \ldots, x_n)$ is constant for all the classes, we can simply say that

$$P(C_i|x_1, x_2, \ldots, x_n) \propto \left(\prod_{j=1}^{j=n} P(x_j|C_i)\right) .P(C_i) \text{ for } 1 \leq i \leq k$$

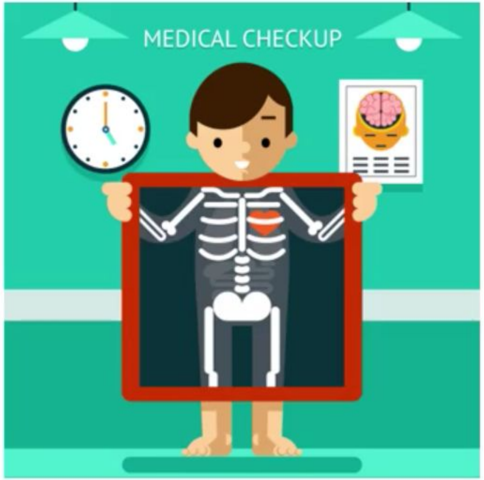
Face Recognition

Weather Prediction



Medical Diagnosis

News Classification



To predict whether a person will purchase a product on a specific combination of **Day**, **Discount** and **Free Delivery** using **Naive Bayes Classifier**

	A	B	C	D
1	Day	Discount	Free Delivery	Purchase
2	Weekday	Yes	Yes	Yes
3	Weekday	Yes	Yes	Yes
4	Weekday	No	No	No
5	Holiday	Yes	Yes	Yes
6	Weekend	Yes	Yes	Yes
7	Holiday	No	No	No
8	Weekend	Yes	No	Yes
9	Weekday	Yes	Yes	Yes
10	Weekend	Yes	Yes	Yes
11	Holiday	Yes	Yes	Yes
12	Holiday	No	Yes	Yes
13	Holiday	No	No	No
14	Weekend	Yes	Yes	Yes
15	Holiday	Yes	Yes	Yes

Naive_Bayes_Dataset

Frequency Table		Buy	
		Yes	No
Discount	Yes	19	1
	No	5	5

Frequency Table		Buy	
		Yes	No
Free Delivery	Yes	21	2
	No	3	4

Frequency Table		Buy	
		Yes	No
Day	Weekday	9	2
	Weekend	7	1
	Holiday	8	3

A

B

FOR OUR BAYES THEOREM, LET THE EVENT BUY BE A AND THE INDEPENDENT VARIABLES, DISCOUNT, FREE DELIVERY AND DAY BE B

Frequency Table		Buy		
		Yes	No	
Day	Weekday	9	2	11
	Weekend	7	1	8
	Holiday	8	3	11
		24	6	30

Likelihood Table		Buy		
		Yes	No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

Frequency Table		Buy: Yes	Buy: No	
Day	Weekday	9	2	11
	Weekend	7	1	8
	Holiday	8	3	11
		24	6	30

Likelihood Table		Buy: Yes	Buy: No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

P(B) = **P(Weekday)**
= 11/30 = 0.37

P(A) = **P(No Buy)**
= 6/30 = 0.2

P(B|A)
= **P(Weekday | No Buy)**
= 2/6 = 0.33

P (A|B) = **P (No Buy | Weekday)**

= P(Weekday| No Buy) * P(No Buy) / P(Weekday)

= (0.33 * 0.2) / 0.367 = 0.179

Frequency Table		Buy		
		Yes	No	
Day	Weekday	9	2	11
	Weekend	7	1	8
	Holiday	8	3	11
		24	6	30

Likelihood Table		Buy		
		Yes	No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

P(B) = **P(Weekday)** = 11/30 = 0.367

P(A) = **P(Buy)** = 24/30 = 0.8

P(B|A) = **P(Weekday | Buy)** = 2/6 = 0.375

If A equals Buy, then

P (A|B) = **P (Buy | Weekday)**

= P(Weekday| Buy) * P(Buy) / P(Weekday)

= (0.375 * 0.8) / 0.367 = 0.817

$P(A|B) = P(\text{No Buy} \mid \text{Weekday})$

$= P(\text{Weekday} \mid \text{No Buy}) * P(\text{No Buy}) / P(\text{Weekday})$

$= (0.33 * 0.2) / 0.367 = 0.179$

$P(B) = P(\text{Weekday}) = 11/30 = 0.367$

$P(A) = P(\text{Buy}) = 24/30 = 0.8$

$P(B|A) = P(\text{Weekday} \mid \text{Buy}) = 2/6 = 0.375$

If A equals Buy, then

$P(A|B) = P(\text{Buy} \mid \text{Weekday})$

$= P(\text{Weekday} \mid \text{Buy}) * P(\text{Buy}) / P(\text{Weekday})$

$= (0.375 * 0.8) / 0.367 = 0.817$

As the **Probability(Buy | Weekday)** is more than **Probability(No Buy | Weekday)**, we can conclude that a customer will most likely buy the product on a Weekday

Likelihood Table		Buy: Yes	Buy: No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

Frequency Table		Buy: Yes	Buy: No	
Discount	Yes	19/24	1/6	20/30
	No	5/24	5/6	10/30
		24/30	6/30	

Frequency Table		Buy: Yes	Buy: No	
Free Delivery	Yes	21/24	2/6	23/30
	No	3/24	4/6	7/30
		24/30	6/30	

Calculating Conditional Probability of purchase on the following combination of day, discount and free delivery:

Where B equals:

- Day = Holiday
- Discount = Yes
- Free Delivery = Yes

Let A = No Buy

P (A|B) = P (No Buy | Discount = Yes, Free Delivery = Yes, Day = Holiday)

$$= \frac{P(\text{Discount = Yes} \mid \text{No}) * P(\text{Free Delivery = Yes} \mid \text{No}) * P(\text{Day = Holiday} \mid \text{No}) * P(\text{No Buy})}{P(\text{Discount=Yes}) * P(\text{Free Delivery=Yes}) * P(\text{Day=Holiday})}$$

$$= \frac{(1/6) * (2/6) * (3/6) * (6/30)}{(20/30) * (23/30) * (11/30)}$$

$$= 0.178$$

Likelihood Table		Buy		
		Yes	No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

Frequency Table		Buy		
		Yes	No	
Discount	Yes	19/24	1/6	20/30
	No	5/24	5/6	10/30
		24/30	6/30	

Frequency Table		Buy		
		Yes	No	
Free Delivery	Yes	21/24	2/6	23/30
	No	3/24	4/6	7/30
		24/30	6/30	

Calculating Conditional Probability of purchase on the following combination of day, discount and free delivery:

Where B equals:

- Day = Holiday
- Discount = Yes
- Free Delivery = Yes

Let A = Buy

P (A|B) = P (Yes Buy | Discount = Yes, Free Delivery = Yes, Day = Holiday)

$$= \frac{P(\text{Discount} = \text{Yes} \mid \text{Yes}) * P(\text{Free Delivery} = \text{Yes} \mid \text{Yes}) * P(\text{Day} = \text{Holiday} \mid \text{Yes}) * P(\text{Yes Buy})}{P(\text{Discount}=\text{Yes}) * P(\text{Free Delivery}=\text{Yes}) * P(\text{Day}=\text{Holiday})}$$

$$= \frac{(19/24) * (21/24) * (8/24) * (24/30)}{(20/30) * (23/30) * (11/30)}$$

$$= 0.986$$

AS **84.71%** IS GREATER THAN **15.29%**, WE CAN CONCLUDE THAT AN AVERAGE CUSTOMER WILL **BUY** ON A **HOLIDAY WITH DISCOUNT AND FREE DELIVERY**

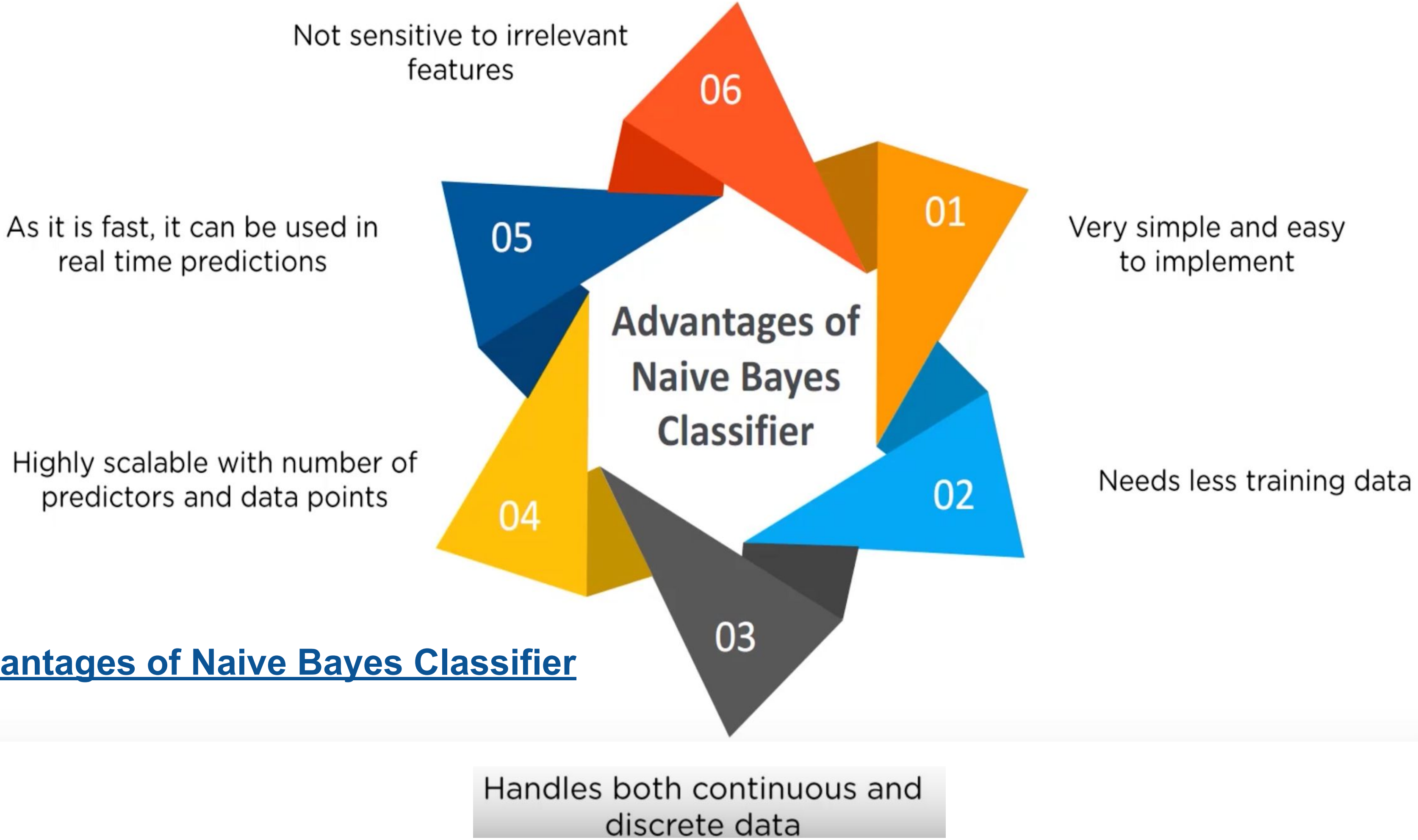
SUM OF PROBABILITIES
= 0.986 + 0.178 = 1.164

LIKELIHOOD OF **PURCHASE**
= 0.986 / 1.164 = **84.71 %**

LIKELIHOOD OF **NO PURCHASE**
= 0.178 / 1.164 = **15.29 %**

PROBABILITY OF PURCHASE = 0.986
PROBABILITY OF NO PURCHASE = 0.178

Advantages of Naive Bayes Classifier



We were able to correctly classify texts into different groups based on which category they belong to using Naive Bayes Classifier

```
predict_category('Jesus Christ')
```

RELIGION

```
predict_category('Sending load to International Space Station ISS')
```

SPACE

```
predict_category('Suzuki Hayabusa is a very fast motorcycle')
```

MOTORCYCLES

```
predict_category('Audi is better than BMW')
```

AUTOS

```
predict_category('President of India')
```

POLITICS

Weather Data Set - Naive Bayes Solution

Which one is the best predictor ?

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

P(Yes) = 9 / 14
P(No) = 5 / 14

Weather Data Set - Naive Bayes Solution

Frequency Tables

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

		Play Golf	
		Yes	No
Temp.	Hot	2 2/9	2 2/5
	Mild	4 4/9	2 2/5
	Cool	3 3/9	1 1/5

		Play Golf	
		Yes	No
Humidity	High	3 3/9	4 4/9
	Normal	6 6/9	1 1/5

		Play Golf	
		Yes	No
Windy	False	6 6/9	2 2/5
	True	3 3/9	3 3/5

Weather Data Set - Naive Bayes Solution

- Let's assume we have a day with:

  Outlook = Rainy

  Temp = Mild

  Humidity = Normal

  Windy = True

  Likelihood of Yes = P(Outlook=Rainy|Yes)*P(Temp=Mild|Yes)*P(Humidity=Normal|Yes)*P(Windy=True|Yes)*P(Yes) =

  2/9 * 4/9 * 6/9 * 3/9 * 9/14 = 0.014109347

  Likelihood of No = P(Outlook=Rainy|No)* P(Temp=Mild|No)*P(Humidity=Normal|No)*P(Windy=True|No)*P(No) =

  3/5 * 2/5 * 1/5 * 3/5 * 5/14 = 0.010285714

- Now we normalize:

  P(Yes) = 0.014109347/(0.014109347+0.010285714) = 0.578368999

  P(No) = 0.010285714/(0.014109347+0.010285714) = 0.421631001