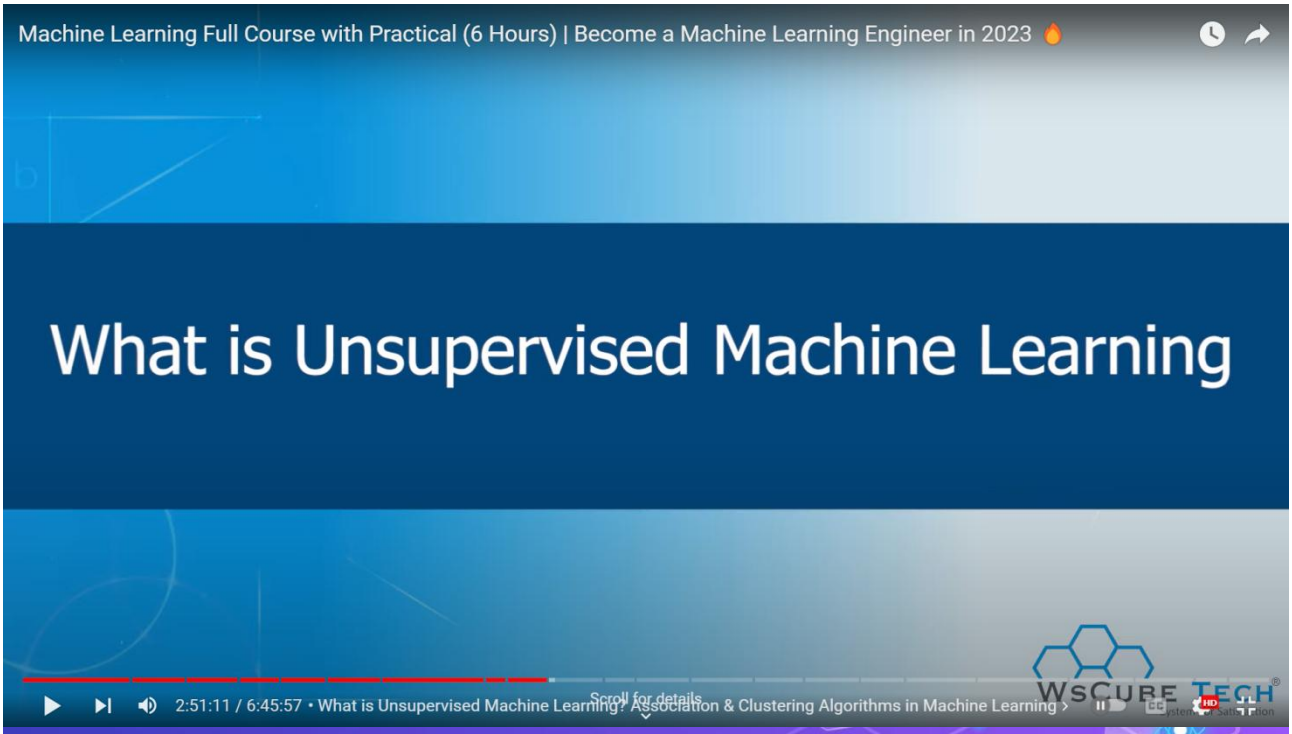
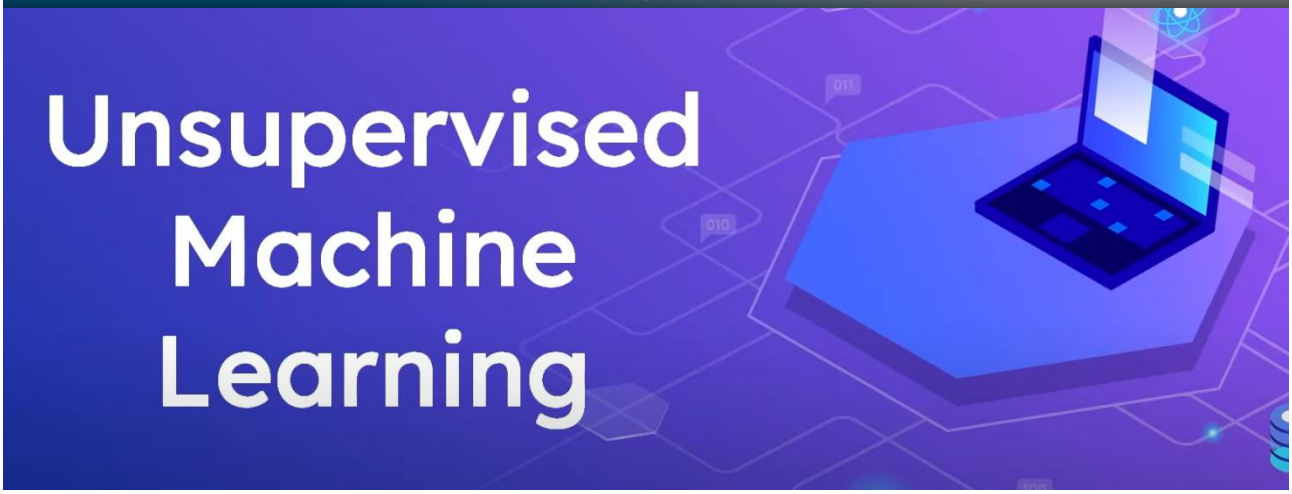
Conclusion

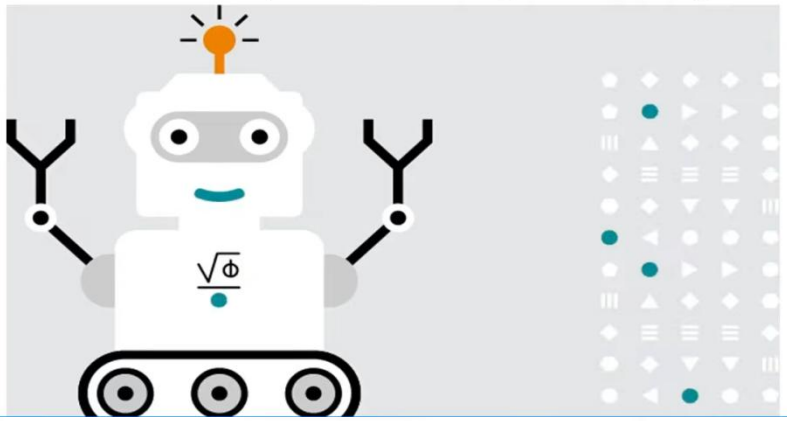
- In Supervised learning, you train the machine using data which is well "labeled."
- Regression and Classification are two types of supervised machine learning techniques.
- The biggest challenge in supervised learning is that if your training set doesn't have examples that you want to have in a class, could give inaccurate results.





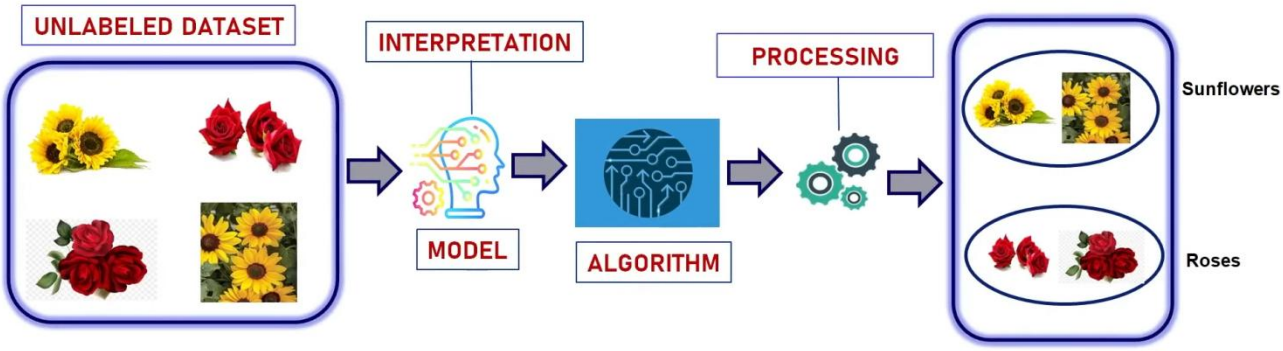
What is Unsupervised Learning?

"In unsupervised learning, you train the machine using data which is "unlabeled" and models itself find the hidden patterns and insights from the given data."



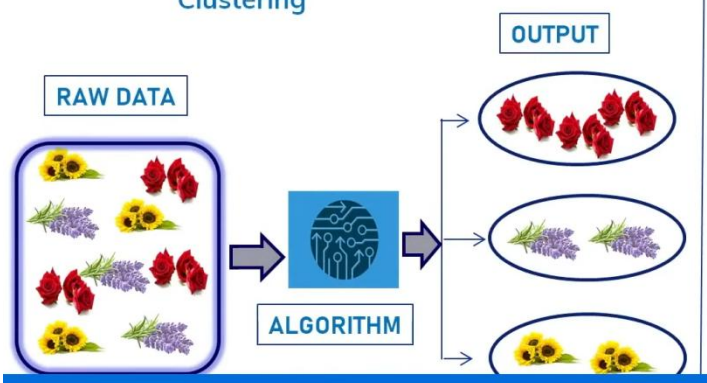
How Does Unsupervised Learning Works?

The goal of Unsupervised Learning is to group unlabelled data according to the similarities, patterns and differences without any prior training of data

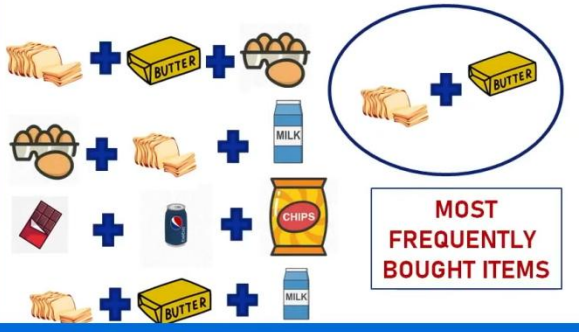


Types of Unsupervised ML Algorithms

Clustering

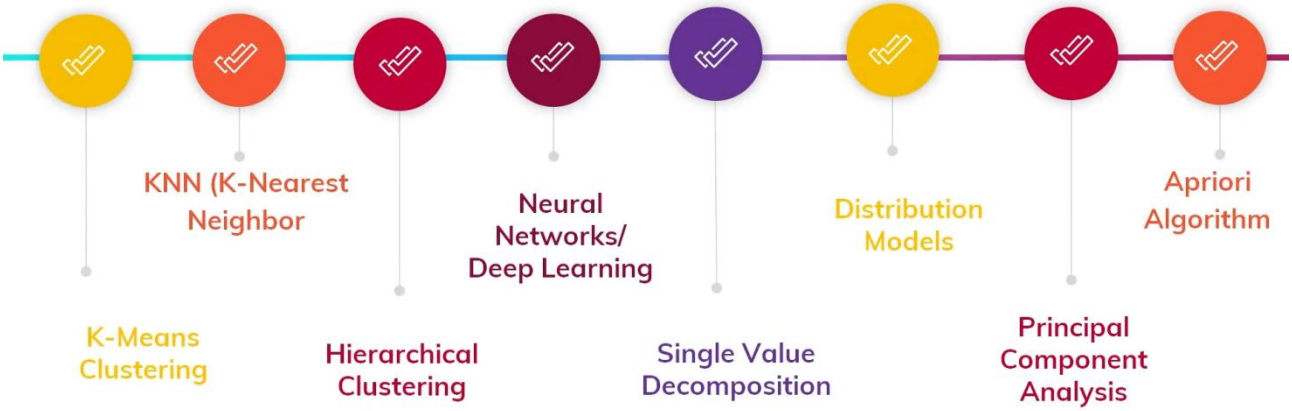


Association



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Unsupervised Learning Algorithms



Advantages of Unsupervised Learning

- Unsupervised learning is used for more complex tasks
- It's helpful in finding patterns in data
- Saves lot of manual work and expense

Disadvantages of Unsupervised Learning

- Less Accuracy
- Time Consuming
- More the features, More the Complexity

Conclusion

- In Unsupervised learning, you train the machine using data which is "unlabeled."
- Clustering and Association are two types of unsupervised machine learning techniques.
- The biggest drawback in unsupervised learning is that it might result in less accuracy.

Training & Testing Data in Machine Learning

Term Insurance Dataset

	Age	Premium
Training Data (80%)	25	18000
	30	32000
	35	42000
	40	47000
Testing Data (20%)	45	55000

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Time – 3.19

Linear Regression in Machine Learning

What is Linear Regression?

Linear Regression is a machine learning algorithm based on supervised learning.

It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as cost, age, sales, temperature, product price, etc.

What is Linear Regression?

Linear Regression is a machine learning algorithm based on supervised learning.

It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as cost, age, sales, temperature, product price, etc.

- Linear Regression with Single Variable
- Linear Regression with Multiple Variables

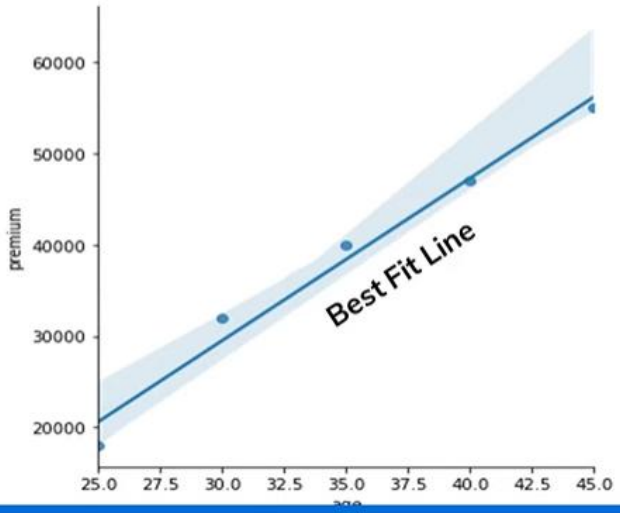
Linear Regression Real World Application

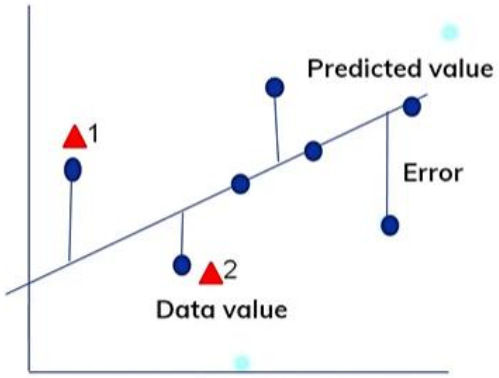
Term Insurance Dataset

Age	Premium
25	18000
30	32000
35	42000
40	47000
45	55000

We'll find out the Premium whose age is 21 and 50

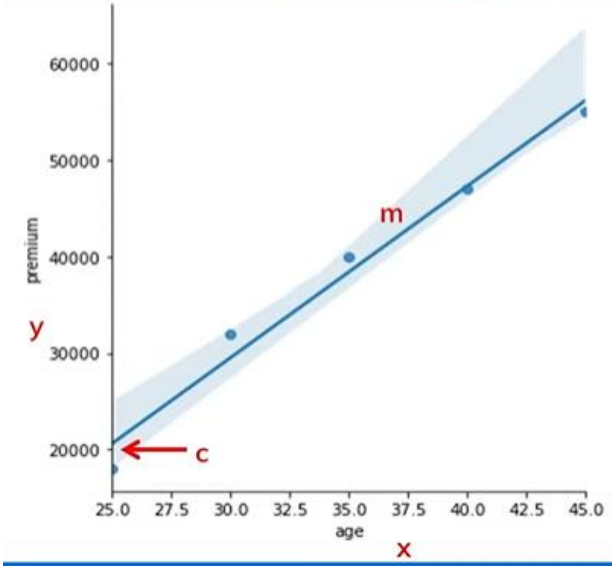
What is Best-Fit Line?





The best fit line will have the least error.

Independent & Dependent Variable



Linear Equation:
y=mx+c

where
y-dependent variable
x –independent variable
m-slope/gradient/coefficient
c-intercept

premium=m*age+c

Go on Jupyter

Validating our Model

Linear Equation: y=mx+c

premium=m*age+c

where
y-dependent variable
x –independent variable
m-slope/gradient/coefficient
c-intercept

Premium whose age is 21

y=m*x+c
= 1780 * 21 + (-23900)
= 37,380 - 23900
13480

Premium whose age is 50

y=m*x+c
= 1780 * 50 + (-23900)
= 89,000 - 23900
65100

Linear Regression Multiple Variables

What is Linear Regression?

Linear Regression is a machine learning algorithm based on supervised learning.

It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as cost, age, sales, temperature, product price, etc.

- Linear Regression with Single Variable
- Linear Regression with Multiple Variables

Linear Regression Real World Application

Term Insurance Dataset

Age	height	weight	Premium
25	162.56	70	18000
30	172.72	95	38000
35	167.64	78	38000
40		110	60000
45	157.48	85	70000

We'll find out the Premium whose age is 27 height is 165.56 and weight is 60 whose age is 60, height is 165.10 and weight is 80

Formulae

Linear Equation: $y = m_1*x_1 + m_2*x_2 + m_3*x_3 + c$

$$premium = m_1*age + m_2*height + m_3*weight + c$$

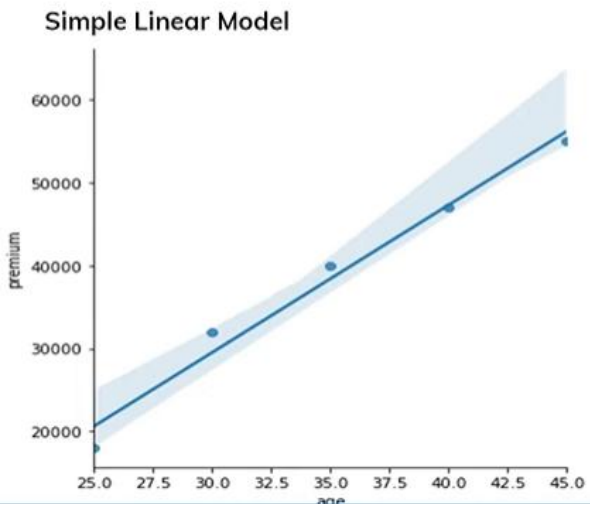
Where,

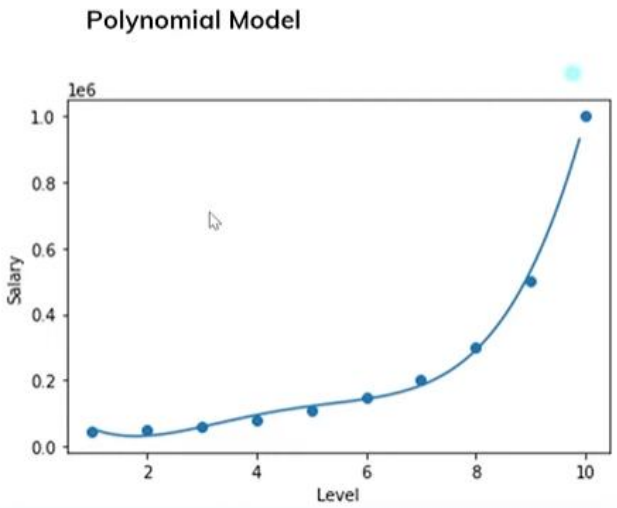
y- dependent variable (premium)
x1 – independent variable (age)
x2 – independent variable (height)
x3 – independent variable (weight)
m1, m2, m3- slope/gradient/coefficient
c-intercept

Time – 3.57

Machine Learning Polynomial Regression Explained

What is Polynomial Regression?





Degree of Polynomial

Linear Equation with Single Variable $y=m_1 x_1 + c$

Linear Equation with Multiple Variables $y = m_1*x_1 + m_2*x_2 + m_3*x_3 + c$

Degree of Polynomial

0 Degree Polynomial	1 Degree Polynomial	2 Degree Polynomial
$y=constant$	$y=mx^1 + c$	$y=ax^2 +bx +c$

Generalized Polynomial Equation

$$y=a_0 + a_1 x + a_2 x^2 + a_3 x^3 + ... + a_n x^n$$

Time – 4.14

Logistic Regression [Binary Classification]

What is Logistics Regression?

Logistic Regression is a machine learning algorithm based on supervised learning.

It is a statistical method that is used for predicting probability of target variable.
Logistic Regression makes probability for classification problems that are discrete in nature.

Example: English or Hindi, True or False, 1 or 0, Right or Wrong, cat or dog or goat.

What is Logistics Regression?

Logistic Regression is a machine learning algorithm based on supervised learning.

It is a statistical method that is used for predicting probability of target variable.
Logistic Regression makes probability for classification problems that are discrete in nature.

Example: English or Hindi, True or False, 1 or 0, Right or Wrong, cat or dog or goat.

- Binary Classification Ex: win or loss, dead or alive
- Multiclass Classification Ex: Onion or Potato or Sweet Potato, Lily or Sunflower or Rose

Logistic Regression Real Life Example

Heart Attack Prediction
Tumour Prediction

Credit-Card Fraud

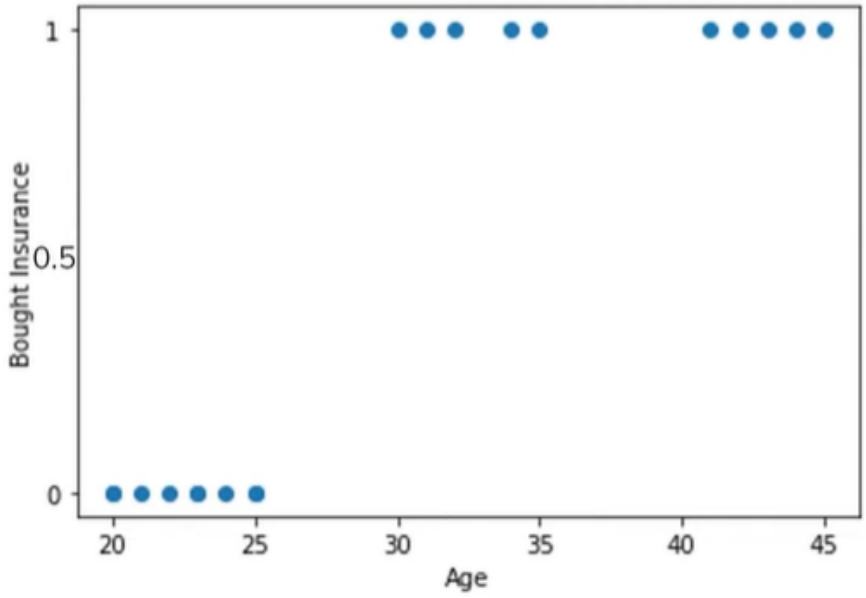
Spam Detection

Term Insurance Dataset

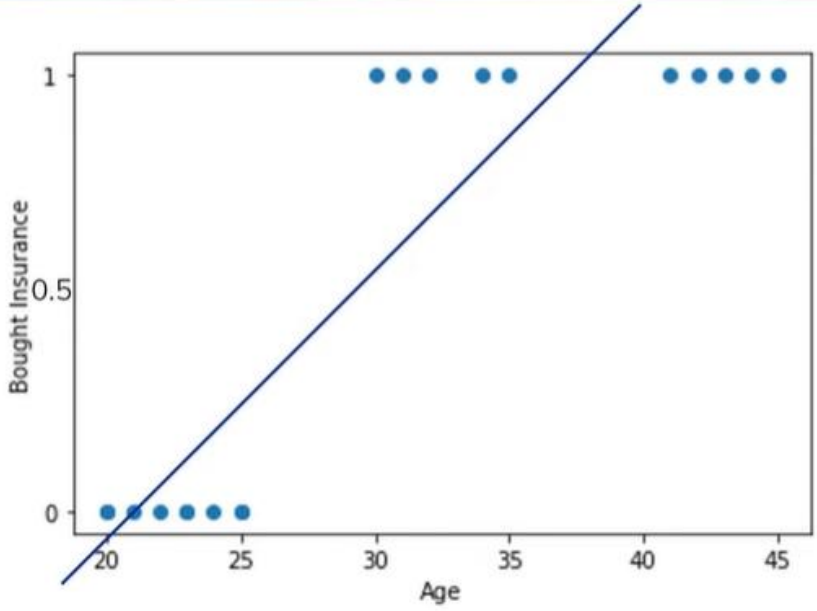
Age	Bought Insurance
21	no
48	yes
32	yes
41	yes
20	no
35	yes
20	no
23	no

Bought Insurance
0
1
1
1
0
1
0
0

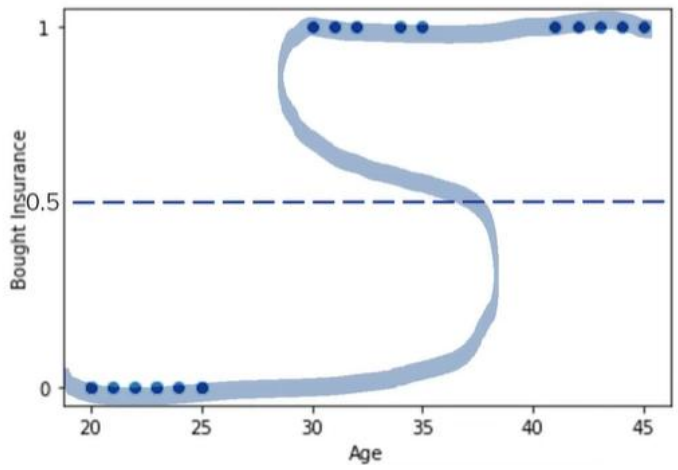
Linear Regression Graph



Linear Regression Graph



Logistic Regression Graph



$$y = 1 / 1+e^{-x}$$

Sigmoid Function
Converts input (x) into range 0 or 1

y - dependent variable (Bought Insurance)

x - independent variable (age)

e - Euler's Constant ~ 2.71828

Time – 4.32

What is Logistics Regression?

Logistic Regression is a machine learning algorithm based on supervised learning.

It is a statistical method that is used for predicting probability of target variable.
Logistic Regression makes probability for classification problems that are discrete in nature.

Example: English or Hindi, True or False, 1 or 0, Right or Wrong, cat or dog or goat.

- Binary Classification Ex: win or loss, dead or alive
- Multiclass Classification Ex: Onion or Potato or Sweet Potato, Lily or Sunflower or Rose

Multiclass Classification Real Life Example

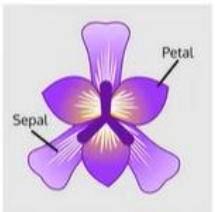
Identity Types of Animal

Identify Type of Vehicle

News Categorization

Iris Dataset

SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
6.8	3.2	5.9	2.3	Iris-virginica
6.9	3.1	5.1	2.3	Iris-virginica
4.9	3.0	1.4	0.2	Iris-setosa
5.6	3.0	4.5	1.5	Iris-versicolor
4.8	3.1	1.6	0.2	Iris-setosa
5.8	2.8	5.1	2.4	Iris-virginica
7.2	3.6	6.1	2.5	Iris-virginica
5.1	3.5	1.4	0.3	Iris-setosa
4.7	3.2	1.6	0.2	Iris-setosa
6.6	3.0	4.4	1.4	Iris-versicolor



Iris Versicolor

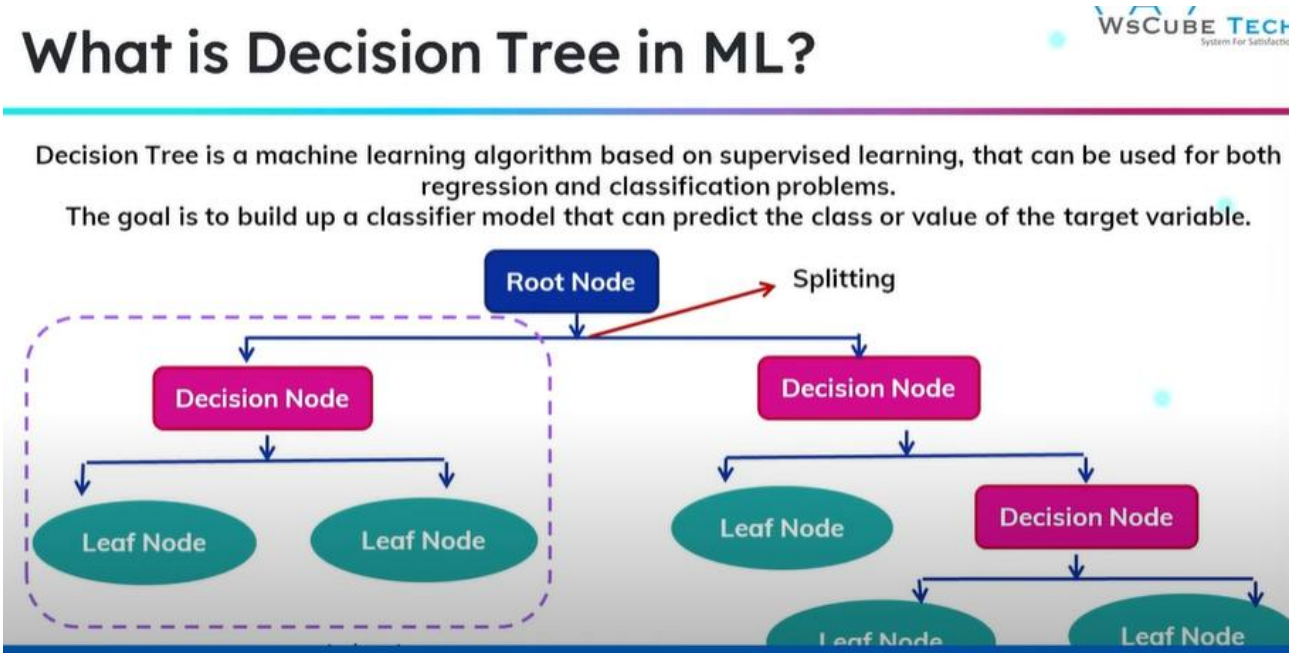
Iris Setosa

Iris Virginica

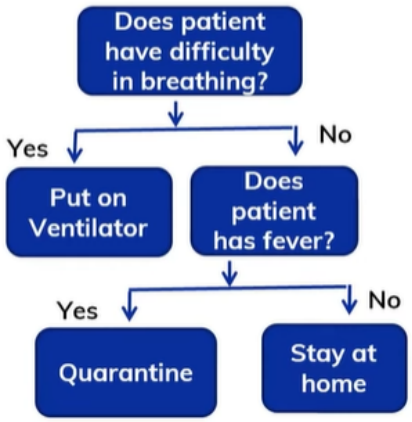
Time- 4.47


Decision Tree
In Machine
Learning

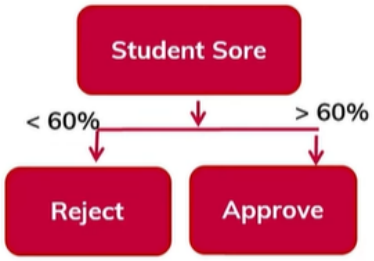
Time – 4.45


What is Decision Tree in ML?
WsCube Tech
Decision Tree is a machine learning algorithm based on supervised learning, that can be used for both regression and classification problems.
The goal is to build up a classifier model that can predict the class or value of the target variable.
Root Node
Splitting
Decision Node
Decision Node
Leaf Node
Leaf Node
Leaf Node
Decision Node
Leaf Node
Leaf Node

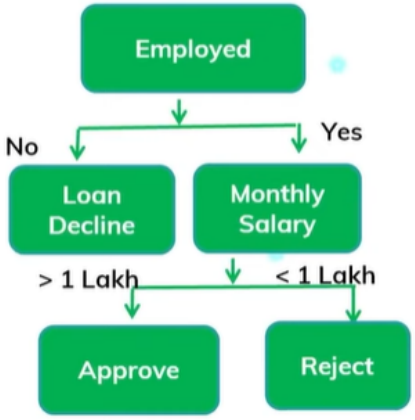
Decision Tree Real Life Example



In Medical Field



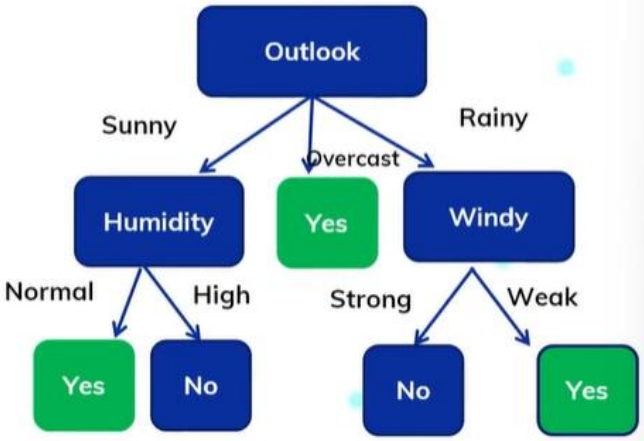
In Educational Sector

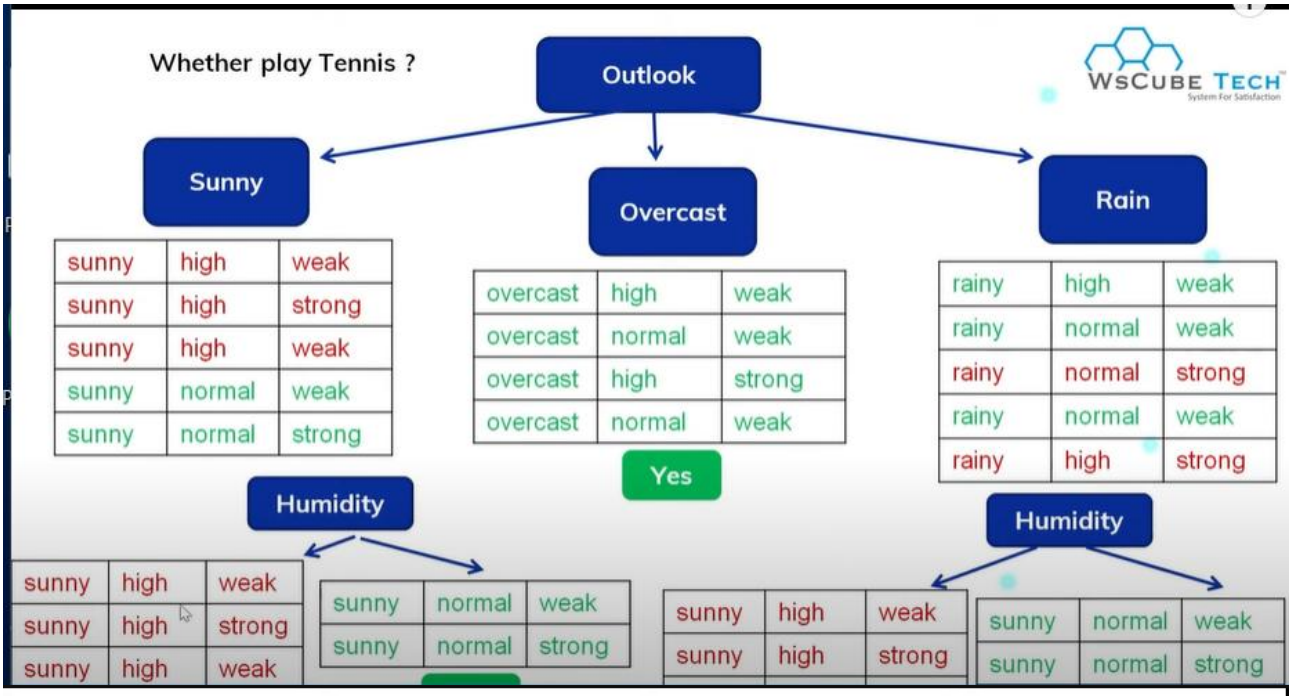


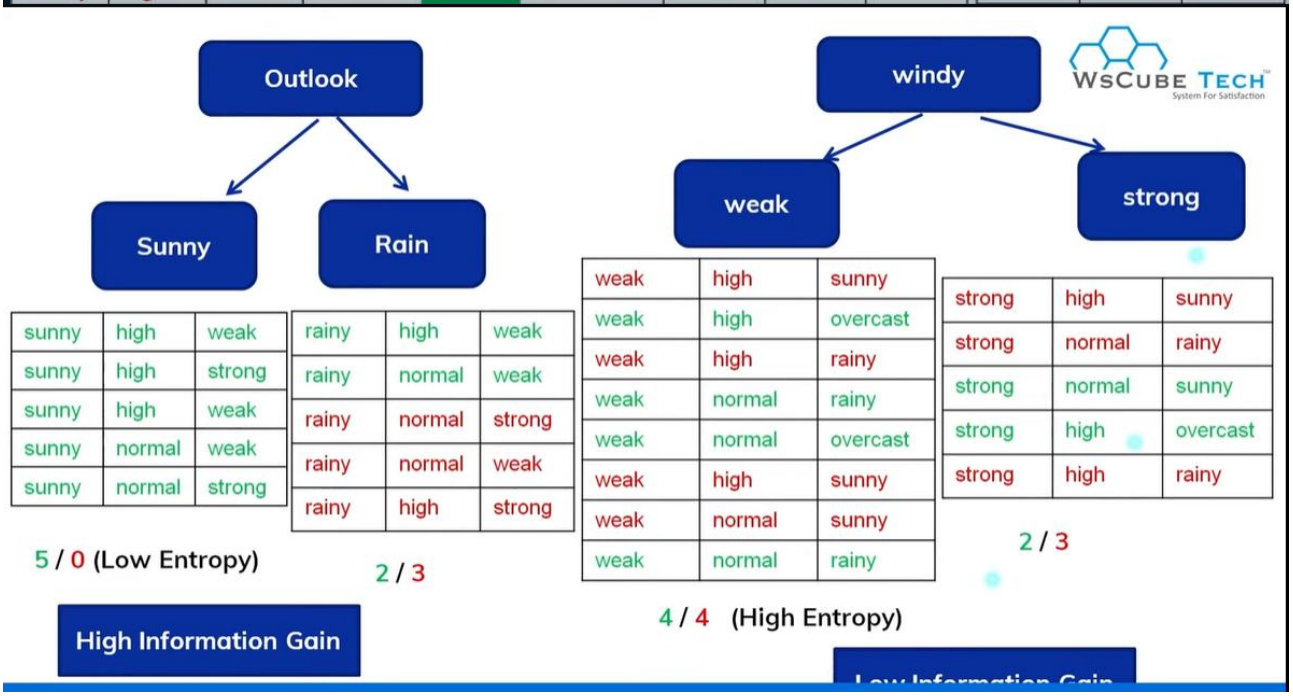
In Banking Sector

How we create a Decision Tree?

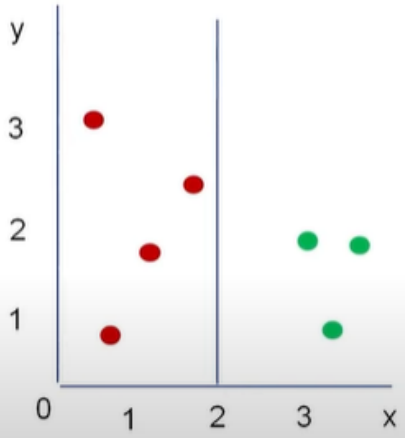
	outlook	humidity	windy	play
1				
2	sunny	high	weak	no
3	sunny	high	strong	no
4	overcast	high	weak	yes
5	rainy	high	weak	yes
6	rainy	normal	weak	yes
7	rainy	normal	strong	no
8	overcast	normal	weak	yes
9	sunny	high	weak	no
10	sunny	normal	weak	yes
11	rainy	normal	weak	yes
12	sunny	normal	strong	yes
13	overcast	high	strong	yes
14	overcast	normal	weak	yes
15	rainy	high	strong	no



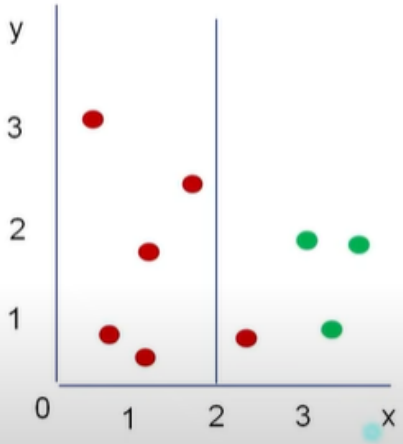

Whether play Tennis ?
Outlook
Sunny
Overcast
Rain
sunny high weak
sunny high strong
sunny high weak
sunny normal weak
sunny normal strong
overcast high weak
overcast normal weak
overcast high strong
overcast normal weak
Yes
rainy high weak
rainy normal weak
rainy normal strong
rainy normal weak
rainy high strong
Humidity
Humidity
sunny high weak
sunny high strong
sunny high weak
sunny normal weak
sunny normal strong
sunny high weak
sunny high strong
sunny normal weak
sunny normal strong
WsCube Tech
System For Satisfaction


Outlook
Sunny
Rain
sunny high weak
sunny high strong
sunny high weak
sunny normal weak
sunny normal strong
rainy high weak
rainy normal weak
rainy normal strong
rainy normal weak
rainy high strong
5 / 0 (Low Entropy)
2 / 3
High Information Gain
windy
weak
strong
weak high sunny
weak high overcast
weak high rainy
weak normal rainy
weak normal overcast
weak high sunny
weak normal sunny
weak normal rainy
strong high sunny
strong normal rainy
strong normal sunny
strong high overcast
strong high rainy
2 / 3
4 / 4 (High Entropy)
WsCube Tech
System For Satisfaction

Gini Impurity



A perfect split



A imperfect split

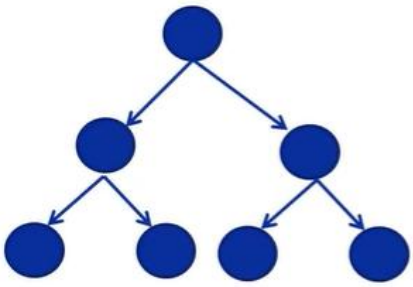
Random Forest Algorithm

What is Random Forest Algorithm ?

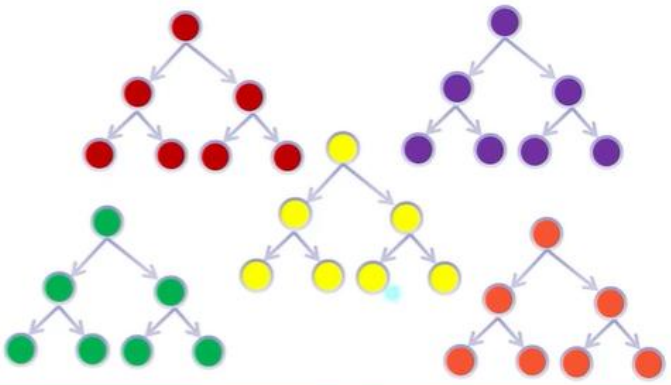
Random Forest is a machine learning algorithm based on supervised learning, that can be used for both regression and classification problems.
It is a collection of multiple random decision trees, which is called the forest.

What is Random Forest Algorithm ?

Random Forest is a machine learning algorithm based on supervised learning, that can be used for both regression and classification problems.
It is a collection of multiple random decision trees, which is called the forest.



Decision Tree

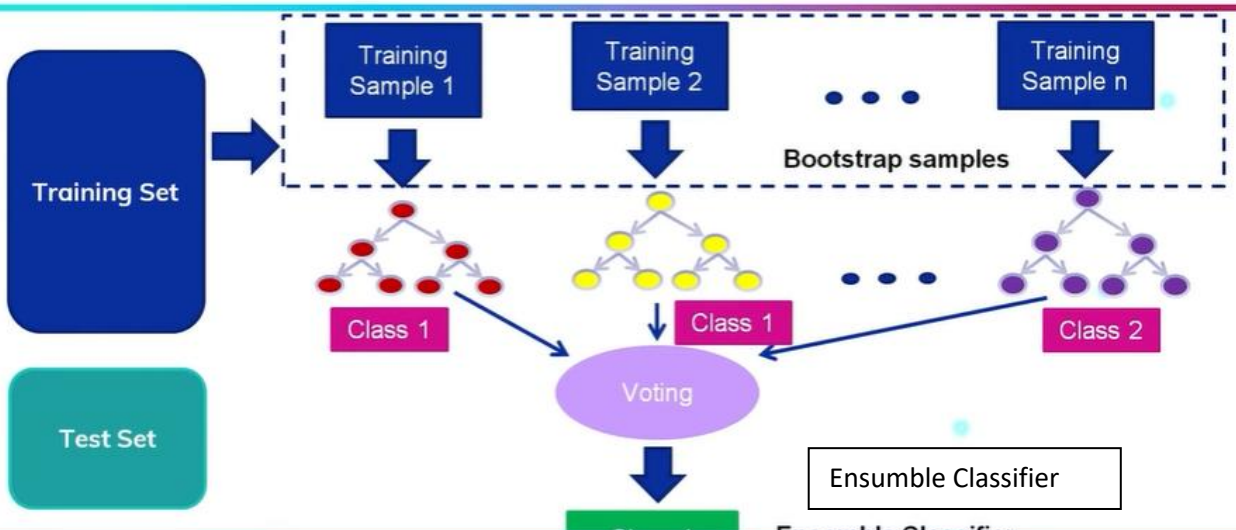


Real Life Example of Random Forest Algorithm

Multiple Decision Trees

Considering opinions of many while

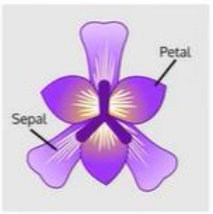
How does the algorithm works ?



Ensumble Classifier

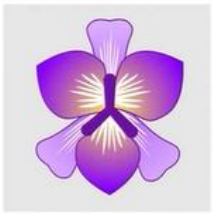
Iris Dataset

SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
6.8	3.2	5.9	2.3	Iris-virginica
6.9	3.1	5.1	2.3	Iris-virginica
4.9	3.0	1.4	0.2	Iris-setosa
5.6	3.0	4.5	1.5	Iris-versicolor
4.8	3.1	1.6	0.2	Iris-setosa
5.8	2.8	5.1	2.4	Iris-virginica
7.2	3.6	6.1	2.5	Iris-virginica
5.1	3.5	1.4	0.3	Iris-setosa
4.7	3.2	1.6	0.2	Iris-setosa
6.6	3.0	4.4	1.4	Iris-versicolor



Iris Versicolor


Iris Setosa


Iris Virginica

TIME – 5.40

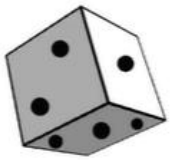
What is Naïve Bayes Classifier Algo ?

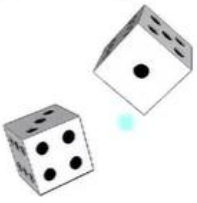
Naïve Bayes Classifier is a machine learning algorithm based on supervised learning, that can be used for solving classification problems.

It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.


Sample Space:
{H,T}


Sample Space:
{HH,HT,TH,TT}


Sample Space:
{1,2,3,4,5,6}


Sample Space:
{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),
(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),
(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),
(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),

Why the Name Naïve Bayes ?

It comprises of two words Naïve and Bayes:

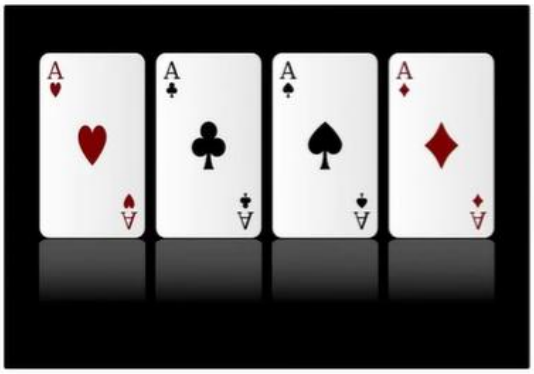
Naïve: It's called naive because it makes the assumption that all attributes are independent of each other.

Bayes: It's called bayes because it depends on the principle of Baye's Theorem

Fruit ={green, oval, sweet} ➡ Watermelon

What is Conditional Probability ?

Playing Card Example



Playing Card Suits

4 Suits = Hearts, Clubs, Diamonds, Spades
13 cards = Ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, King
Each of the 13 cards has all 4 suits.

If you pick a card from the deck, can you guess the probability of getting a king given the card is a spade?

Total spade=13

King=1

P(king/ spade) =1/13 = 0.076

P(A|B)= probability of A occurring given that B has already occurred.

What is Bayes Theorem ?

The Naive Bayes classifier works on the principle of conditional probability, as given by the Bayes theorem.

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Where,

A,B= Events

P(A|B)= Probability of A given that B is true

P(B|A)= Probability of B given that A is true

P(A)= Probability of event A

P(B)= Probability of event B

Bayes Theorem [contd...]

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

$$P(\text{king} \mid \text{spade}) = \frac{P(\text{spade} \mid \text{king}) * P(\text{king})}{P(\text{spade})}$$

$$= \frac{1/4 * 1/13}{1/4} = 1 / 13 = 0.076$$

P (spade / king) =1/4

P (king) = 1/13

P (spade) = 1/4

Where Naïve Bayes is Used?

Face Recognition

Weather Forecast

News Categorization

TIME = 6.02

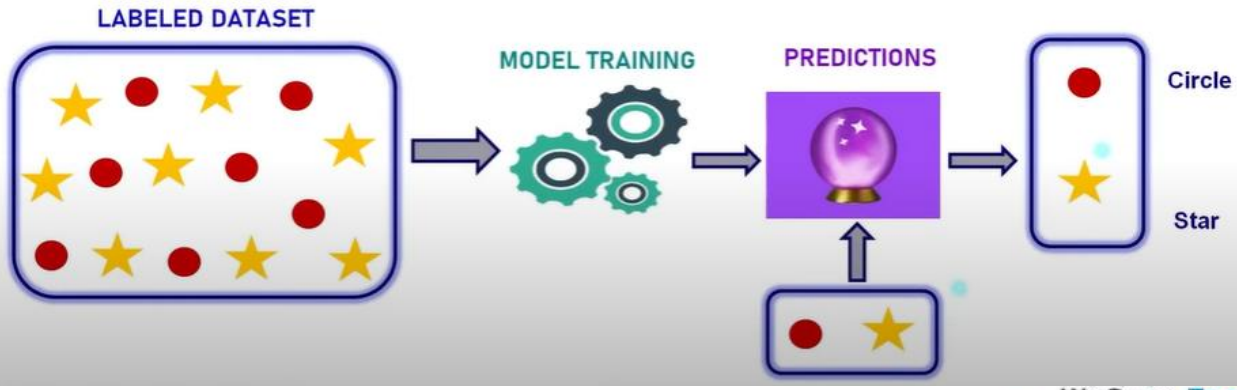
WsCube Tech
System For Satisfaction

Support Vector Machine Algorithm

What is Support Vector Machine ?

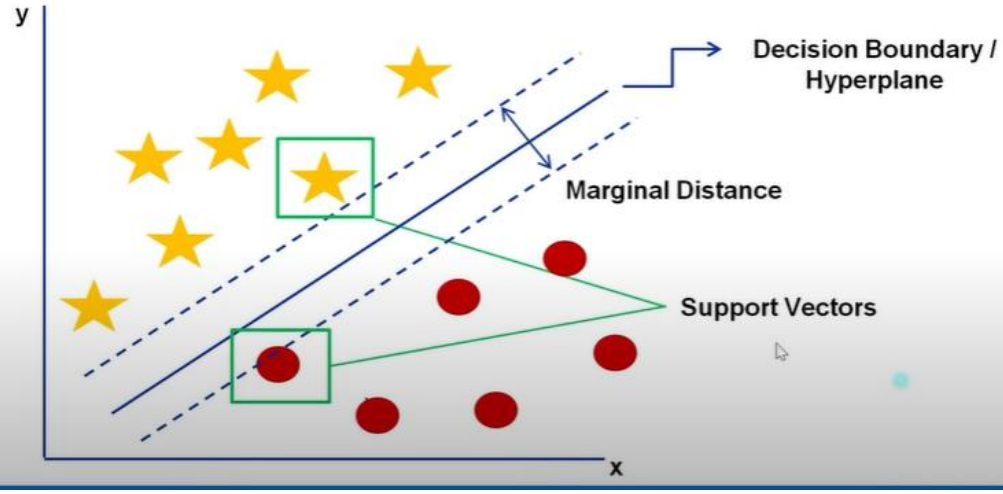
WsCube Tech
System For Satisfaction

Support Vector Machine is a machine learning algorithm based on supervised learning, that can be used for both regression and classification problems.

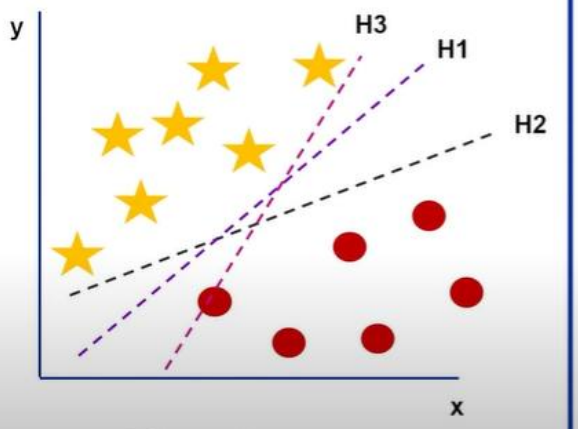


Support Vector Machine Terminologies

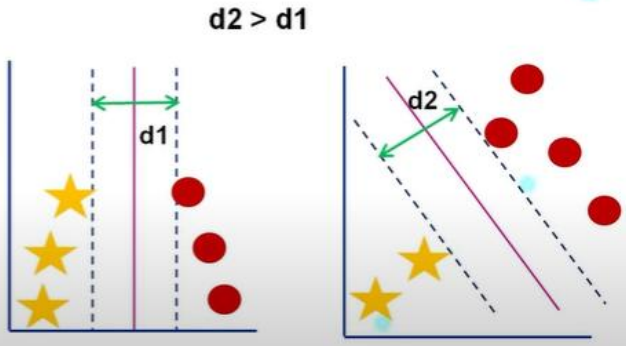
WsCube Tech
System For Satisfaction



Which Hyperplane to Select ?

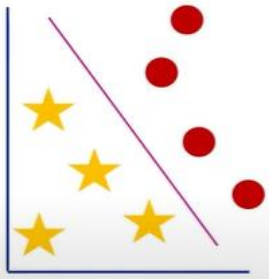


Different Hyperplane
(H1, H2, H3)

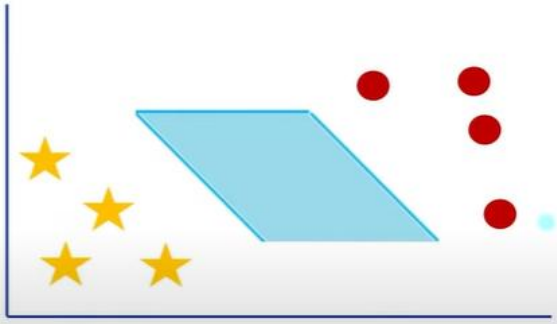


Maximal Margin Hyperplane

Hyperplanes in 2D and 3D Feature Space



2D Space

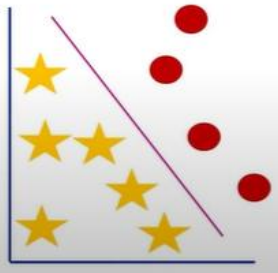


3D Space

Types of SVM

WsCube Tech

Linear SVM is used when dataset can be classified into 2 classes using a straight line



Linear SVM

Non-Linear SVM is used when the dataset cannot be classified into 2 classes using a straight line

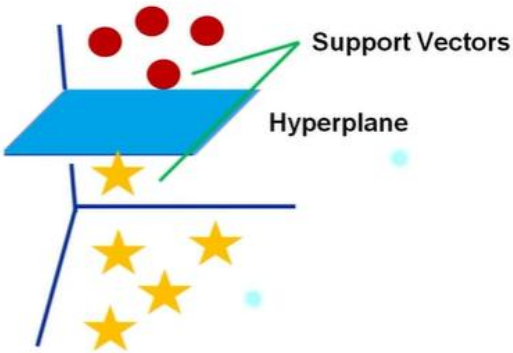
Kernel Function

WsCube Tech

Kernel functions takes low dimensional input space and transform it into a higher-dimensional space, i.e., it converts not separable problem to separable problem.

Low-Dimensional Space

Kernel

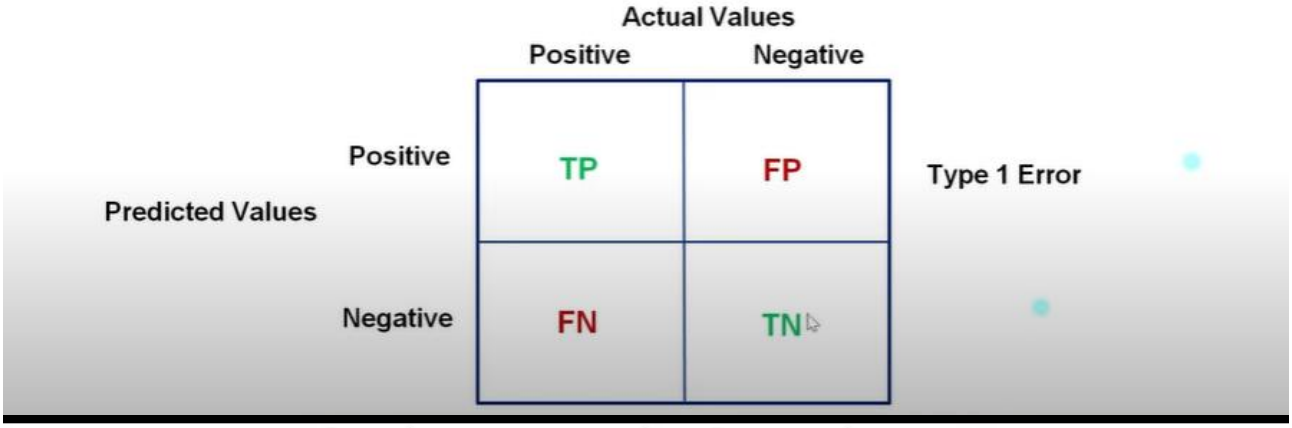


High-Dimensional Space

TIME – 6.22

What is Confusion Matrix ?

A confusion matrix is an N*N matrix used for evaluating the performance of a classification model, where N is the number of target classes.
This compares the actual target values with predicted target values.



Confusion Matrix for Binary Classification Problem

n=200	Predicted: No	Predicted: Yes	
Actual: No	TN=100	FP=15	115
Actual: Yes	FN=5	TP=80	85
	105	95	

True Negatives (TN): The patients actually don't have disease and our model also says that these patients don't have disease

True Positives (TP): These are cases in which, The patients actually have disease and our model also says that these patients do have disease

False Positives (FP): These are cases in which, The patients actually don't have disease but our model says that these patients do have disease

False Negatives (FN): These are cases in which, The patients actually have disease but our model says that these patients don't have disease

Accuracy, Error Rate, Precision, Recall

ACCURACY
The no of predictions that the model got right

Accuracy=(TP + TN) / (TP + TN + FP + FN) = 80+100 / 200 = 0.9

Error Rate
The no of predictions that the model predicted wrongly

Error Rate= (FP + FN) / (TP + TN + FP + FN) = 15 + 5 / 200 = 0.1

Precision
When the model predicts the positives, how often is it right?

Precision = TP / (TP + FP) =80 / 95 =0.86

Recall
When it's actually yes, how often does it predict yes?

Recall = (TP) / (TP + FN) = 80 / 85 = 0.94

n=200	Predicted: No	Predicted: Yes	
Actual: No	TN=100	FP=15	115
Actual: Yes	FN=5	TP=80	85
	105	95	

TIME – 6.32

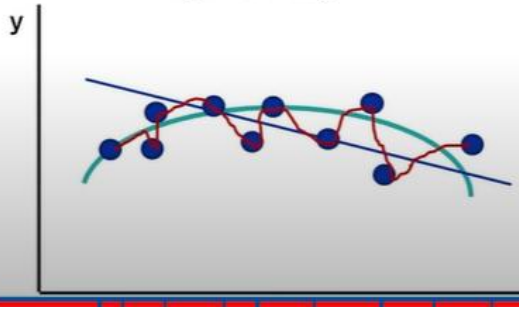
Underfitting and Overfitting

Underfitting and Overfitting

Model's performance is evaluated based upon Accuracy and Generalisation.

Accuracy: It states how well a model predicts the right output
Generalisation: It states how well model behaves on new data set

A model is said to be best when it behaves nearly same way on training as well as test data with high accuracy

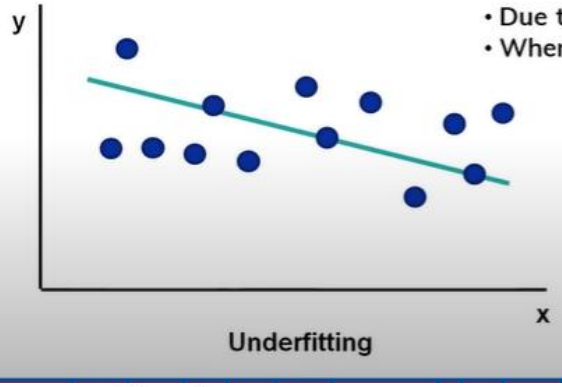


Underfitting

Underfitting means model has low accuracy score on training as well as on test data

Underfitting occurs:

- Due to lack of data
- When we try to build a linear model with a non-linear data.

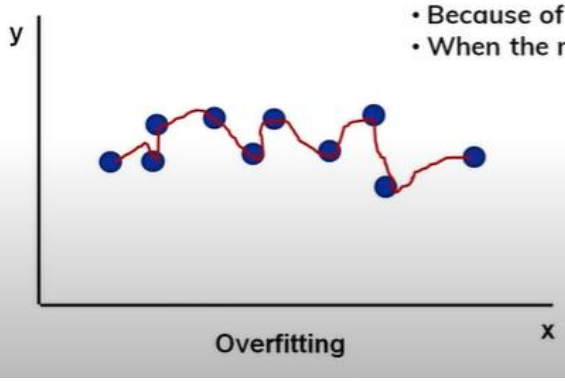


Overfitting

Overfitting means model has high accuracy score on training data but low score on test data

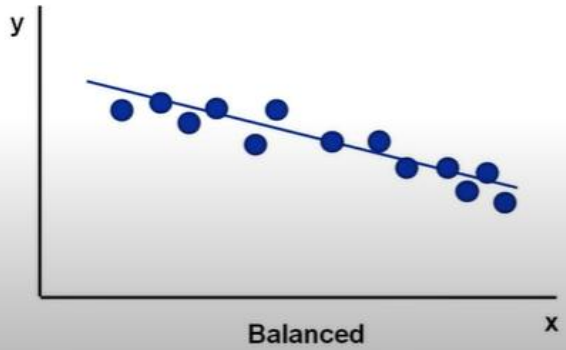
Overfitting occurs:

- Because of a lot more of data
- When the model fits the data too well.



Right Fit

Ideally, the case when the model makes the predictions with 0 error, is said to have a good fit on the data.



Real Life Example of Underfitting

Underfitting Scenario

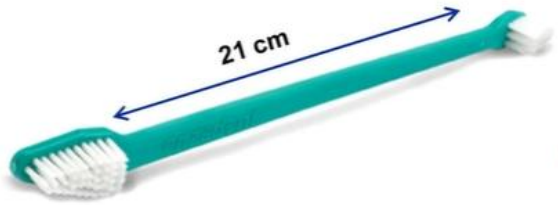
Single Feature:
Bristle size : 0.15 mm

It's a Toothbrush

Real Life Example of Overfitting

Overfitting Scenario

Feature:
bristle size : 0.15 mm
bristle type: soft
used in: mouth
handle : 16 cm



It's Not a Toothbrush

WsCube Tech™
System For Satisfaction

Thanks!

@wscube

@wscubetech.india

www.wscubetech.com

Subscribe To Get Notified about New Videos