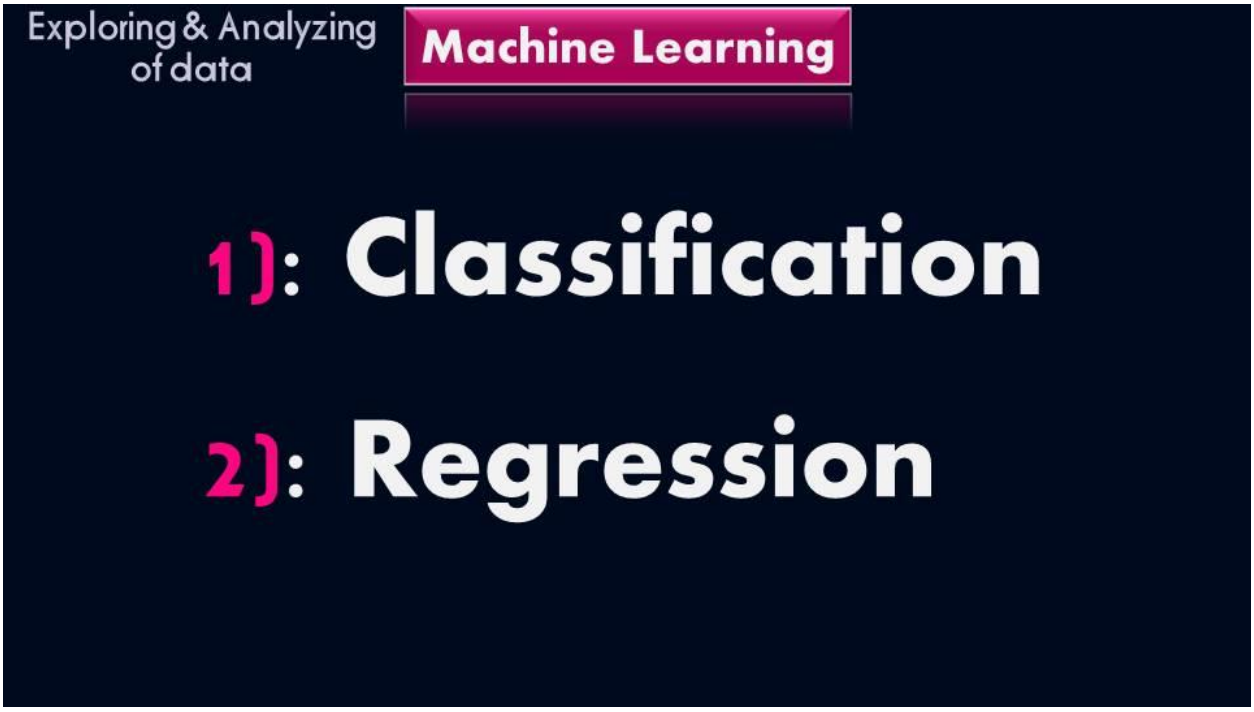
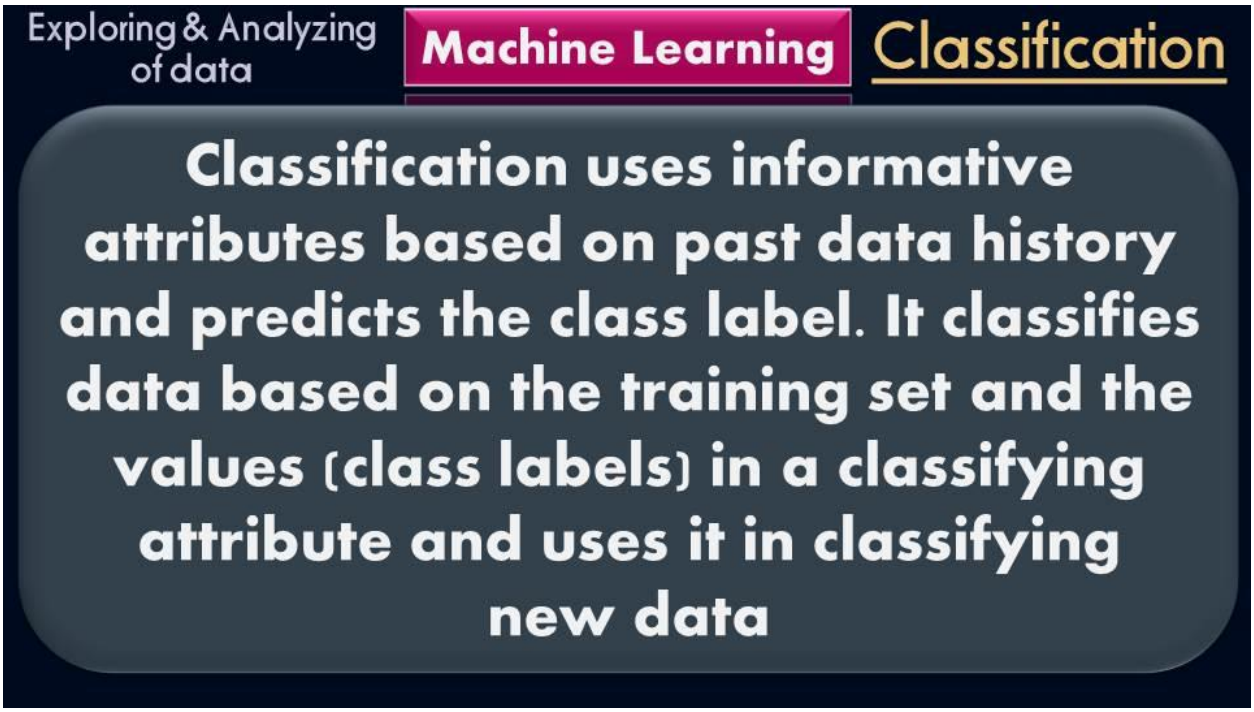
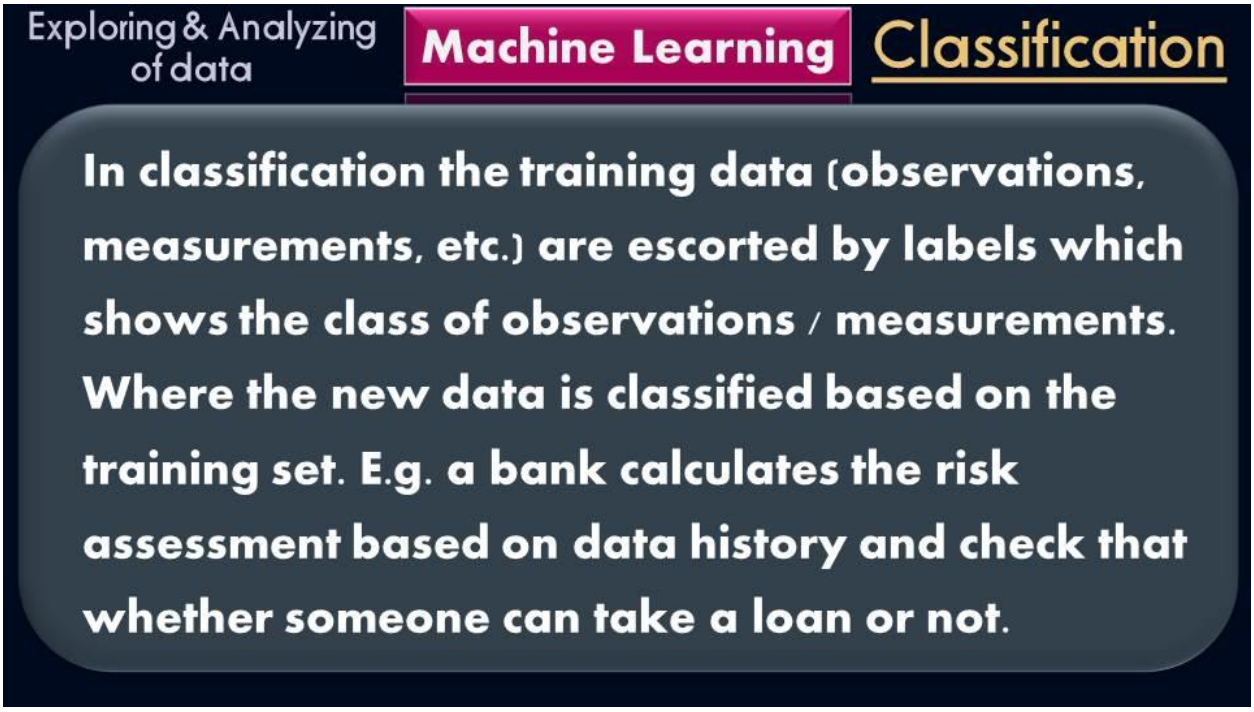
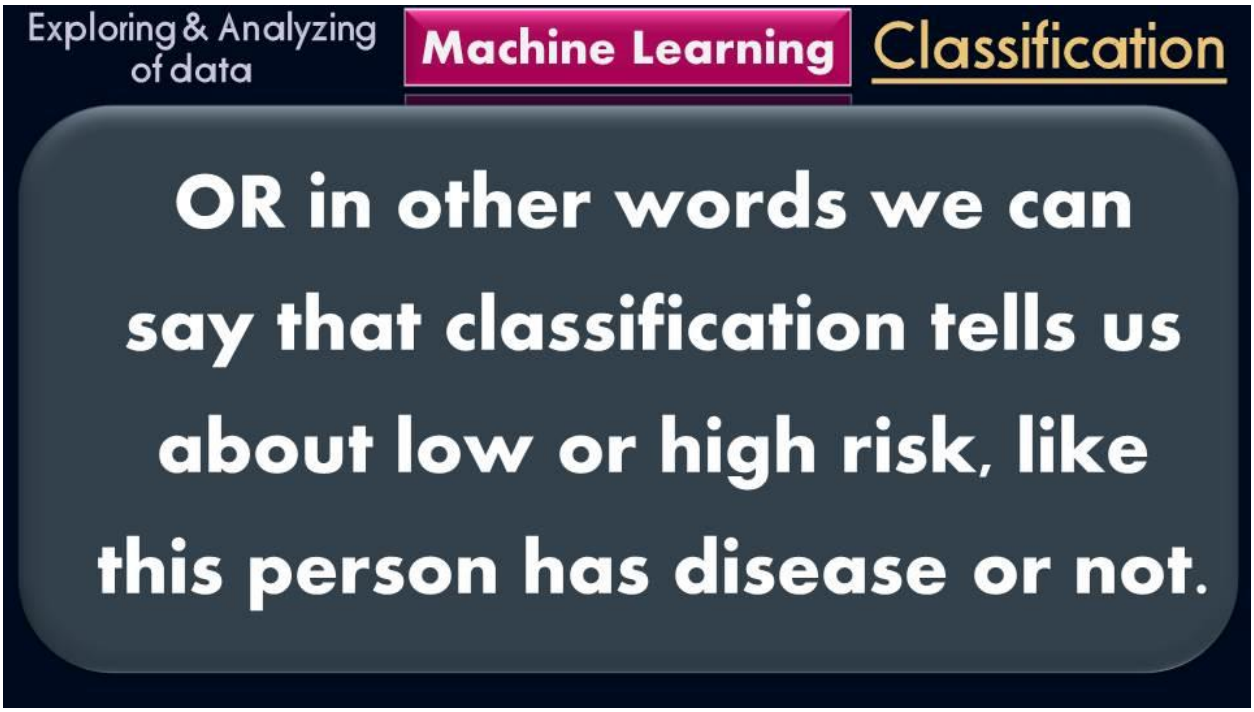
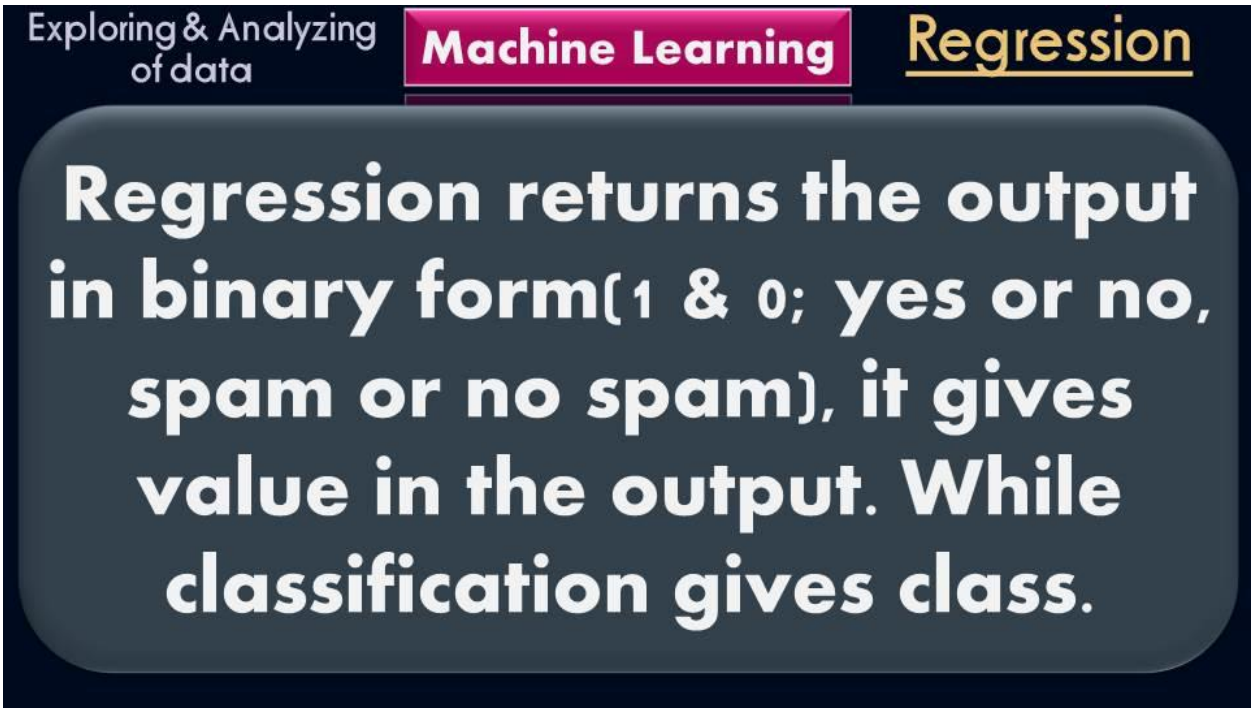

Exploring & Analyzing of data
Machine Learning
1): Classification
2): Regression

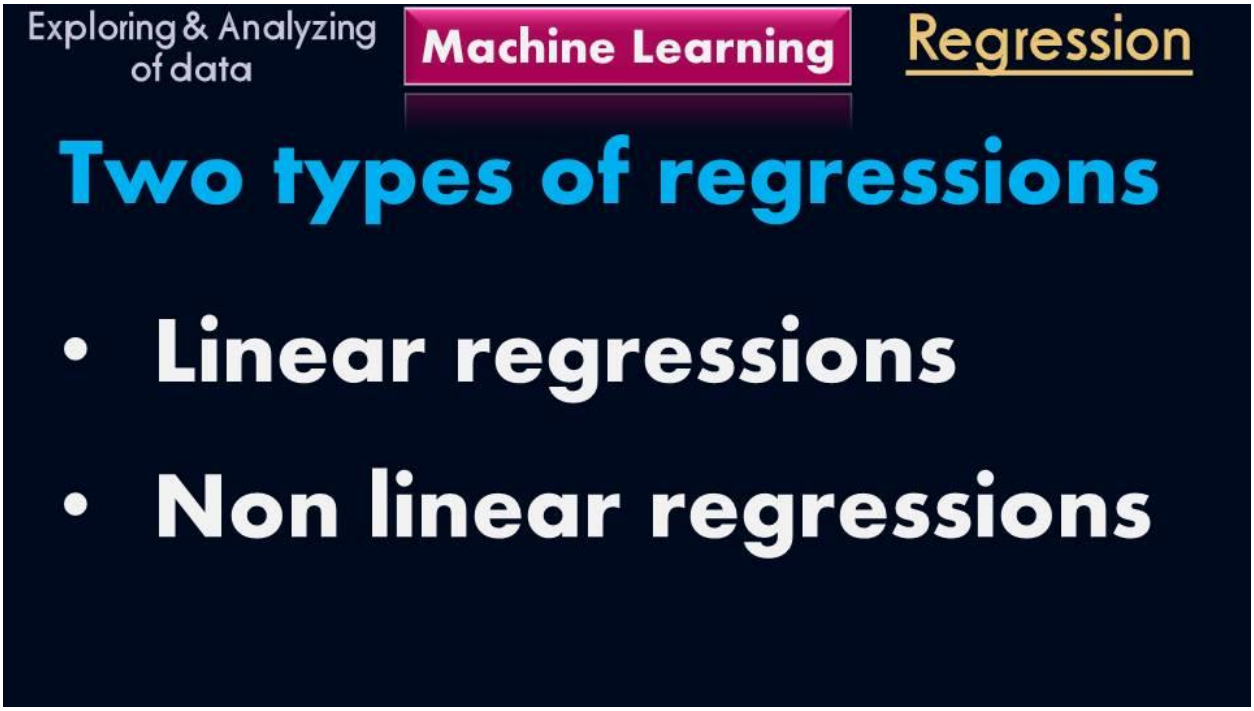

Exploring & Analyzing of data
Machine Learning
Classification
Classification uses informative attributes based on past data history and predicts the class label. It classifies data based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data

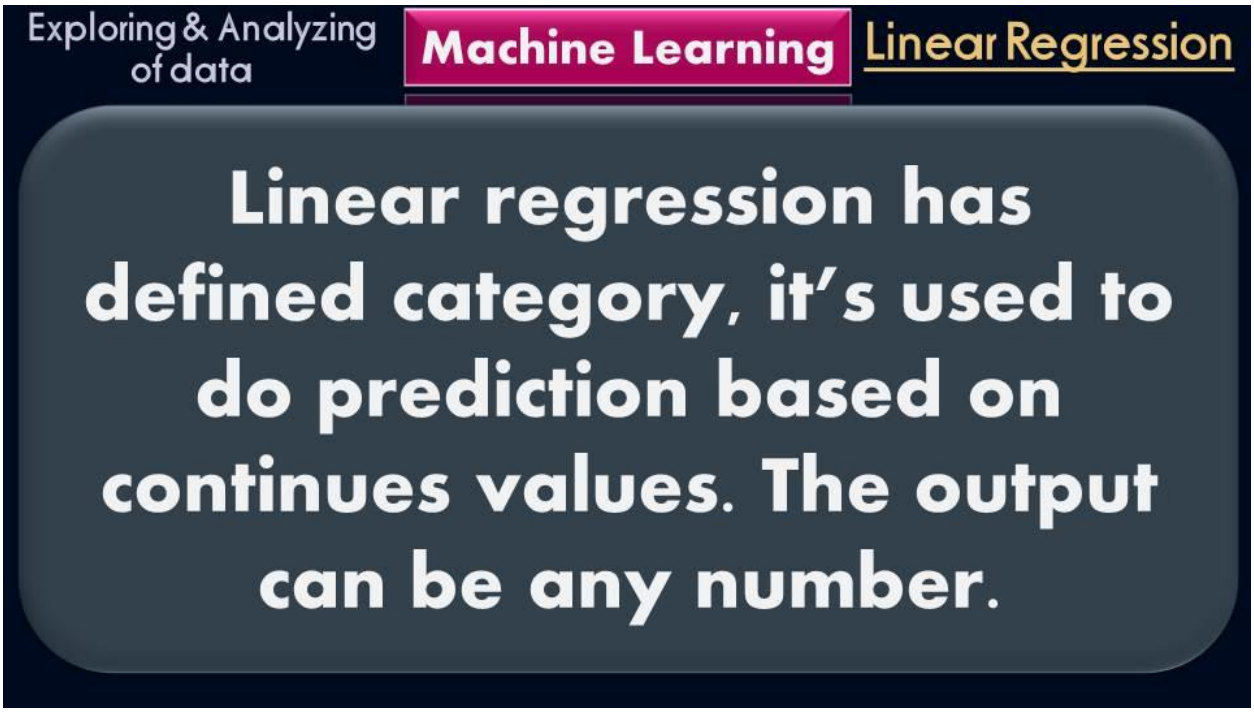

Exploring & Analyzing of data
Machine Learning
Classification
In classification the training data (observations, measurements, etc.) are escorted by labels which shows the class of observations / measurements. Where the new data is classified based on the training set. E.g. a bank calculates the risk assessment based on data history and check that whether someone can take a loan or not.

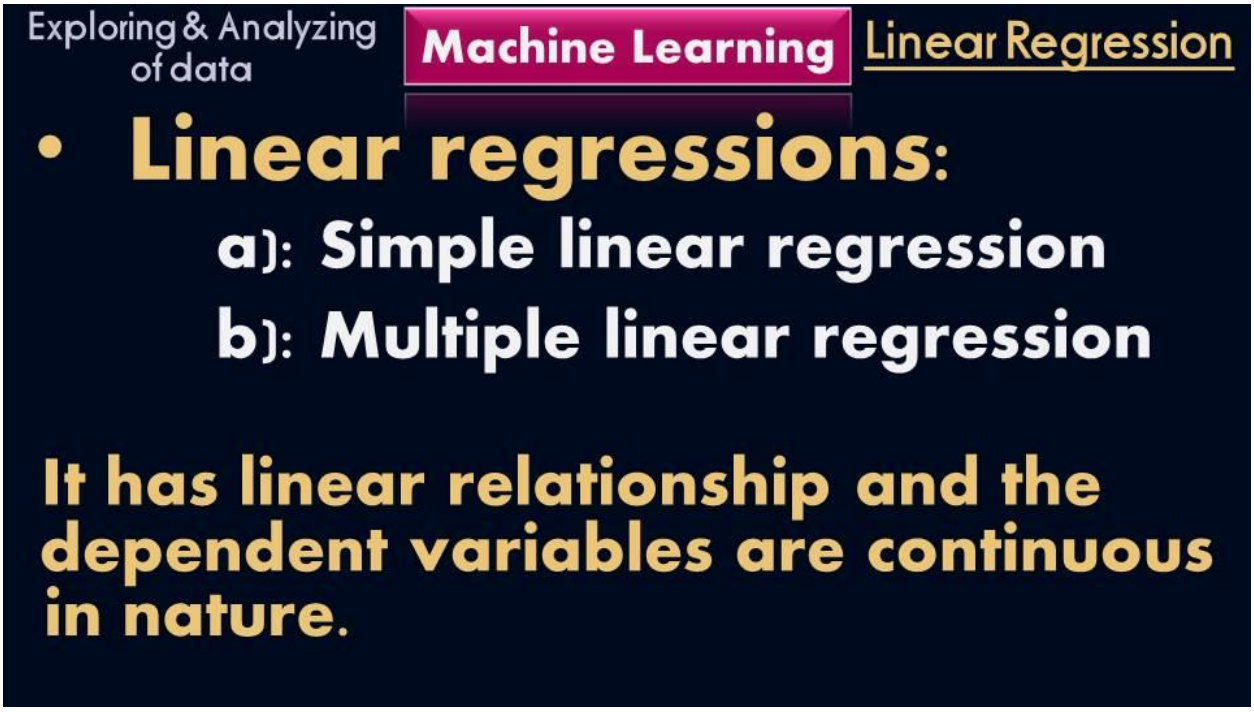

Exploring & Analyzing of data
Machine Learning
Classification
OR in other words we can say that classification tells us about low or high risk, like this person has disease or not.

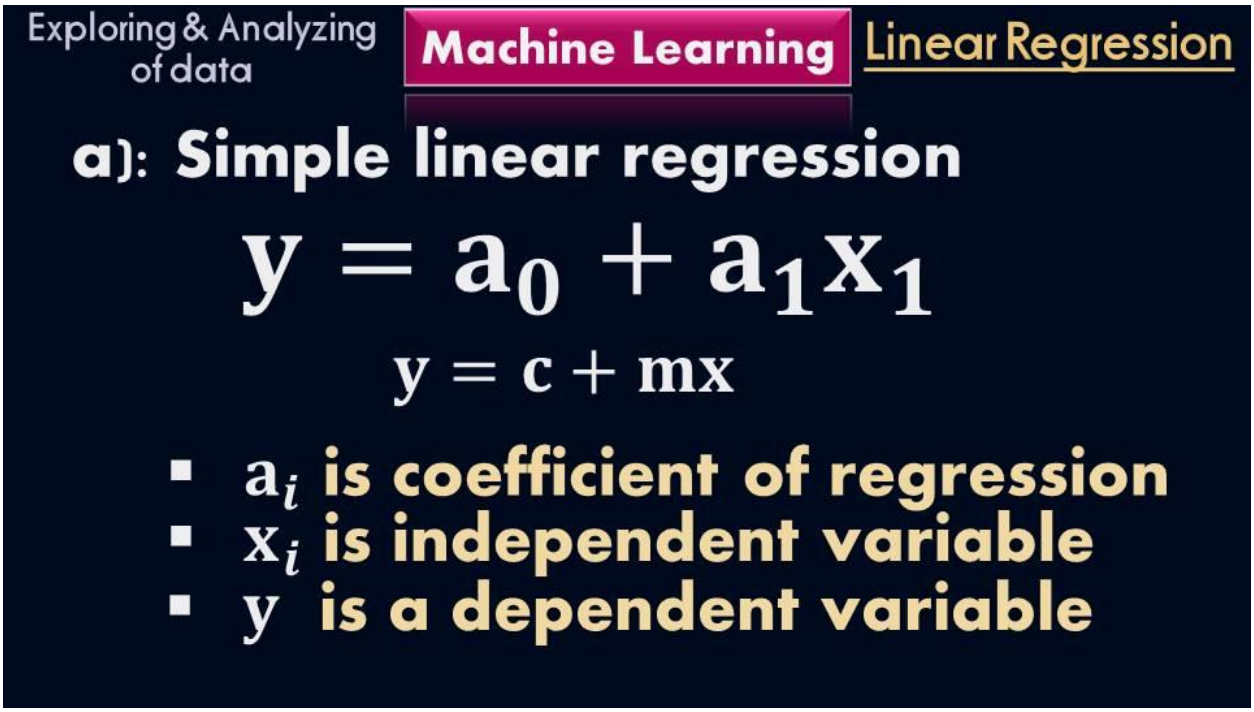

Exploring & Analyzing of data
Machine Learning
Regression
Regression uses continues values to predict next value, like prediction of stock market for tomorrow or weather forecasting.

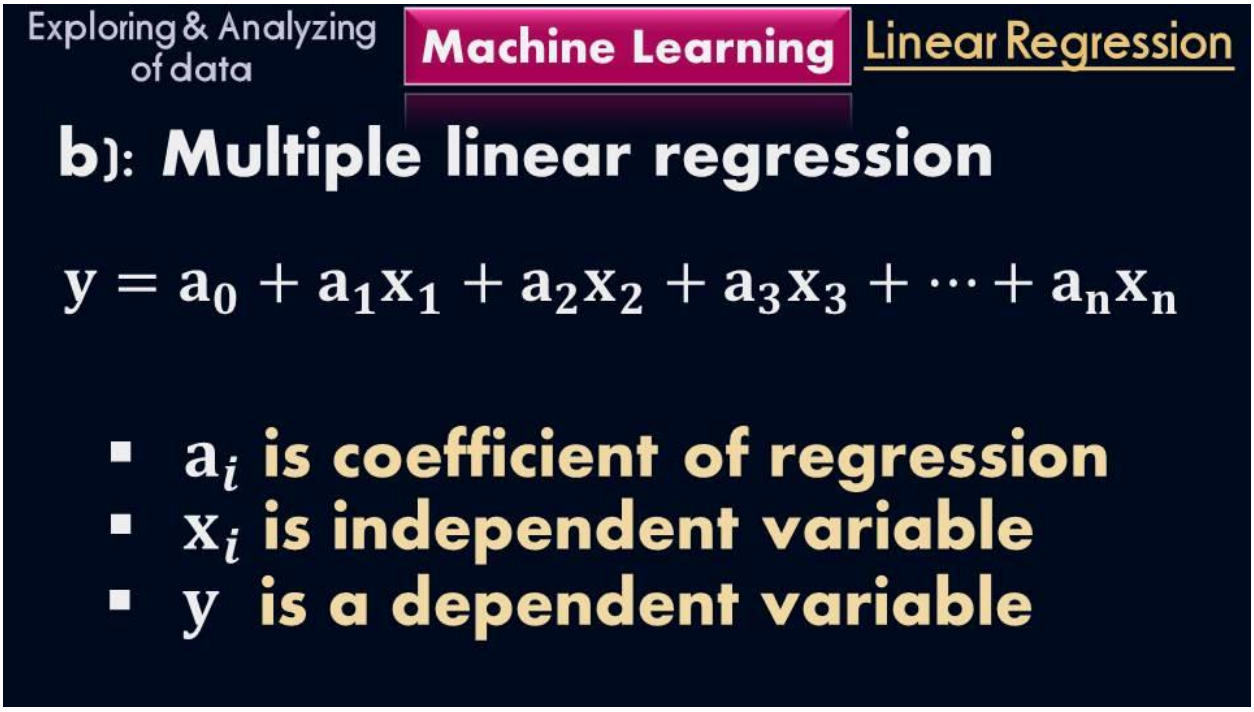

Exploring & Analyzing of data
Machine Learning
Regression
Regression returns the output in binary form(1 & 0; yes or no, spam or no spam), it gives value in the output. While classification gives class.

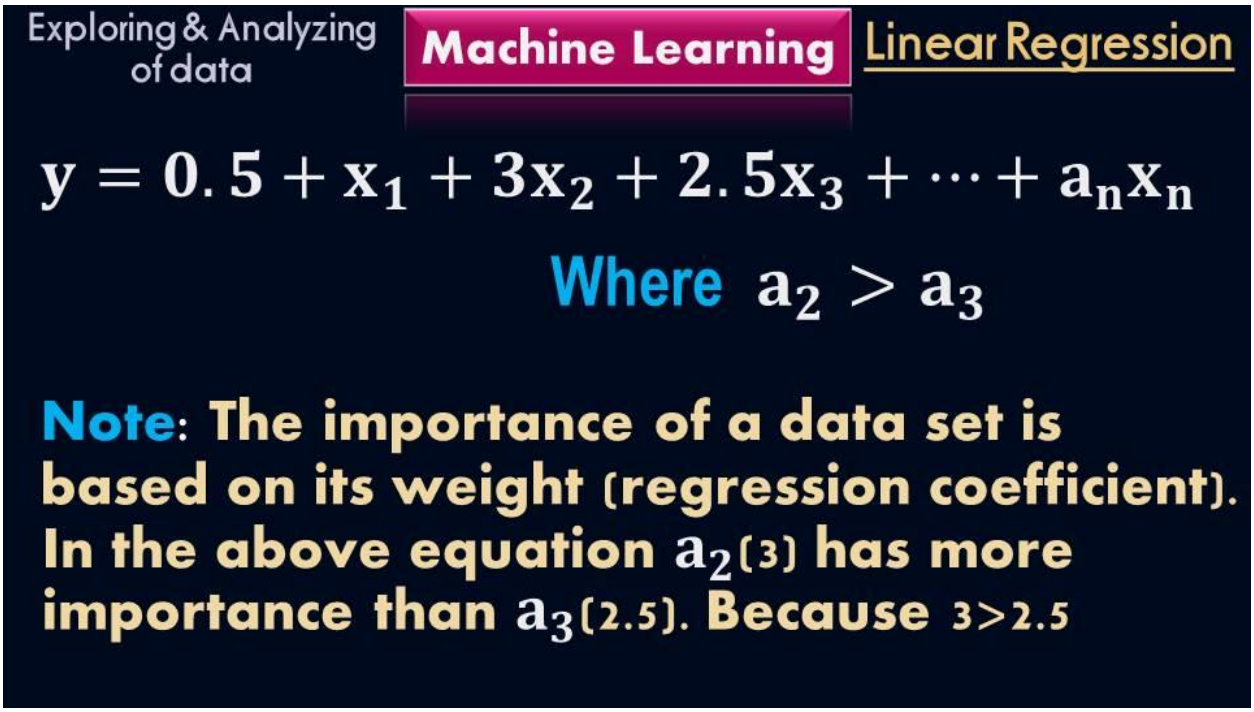

Exploring & Analyzing of data
Machine Learning
Regression
Two types of regressions
• Linear regressions
• Non linear regressions

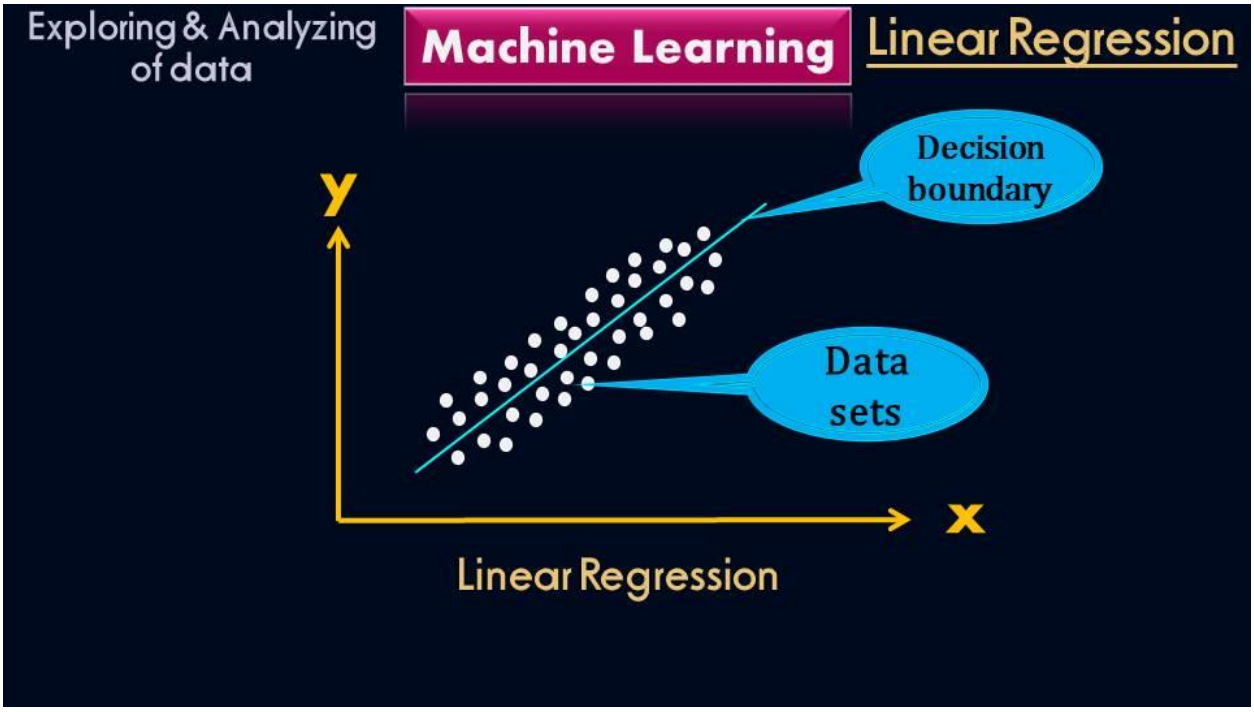

Exploring & Analyzing of data
Machine Learning
Linear Regression
Linear regression has defined category, it's used to do prediction based on continues values. The output can be any number.


Exploring & Analyzing of data
Machine Learning
Linear Regression
• Linear regressions:
a): Simple linear regression
b): Multiple linear regression
It has linear relationship and the dependent variables are continuous in nature.

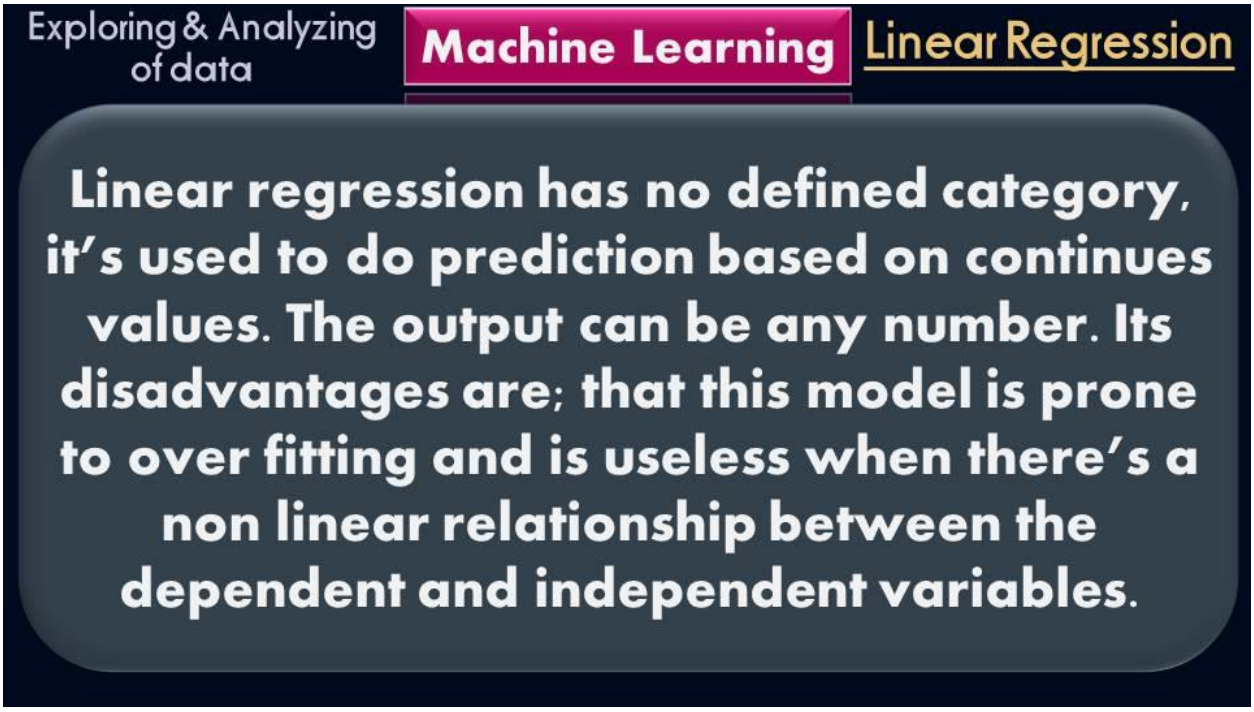

Exploring & Analyzing of data
Machine Learning
Linear Regression
a): Simple linear regression
$y = a_0 + a_1 x_1$
$y = c + mx$
▪ a_i is coefficient of regression
▪ x_i is independent variable
▪ y is a dependent variable

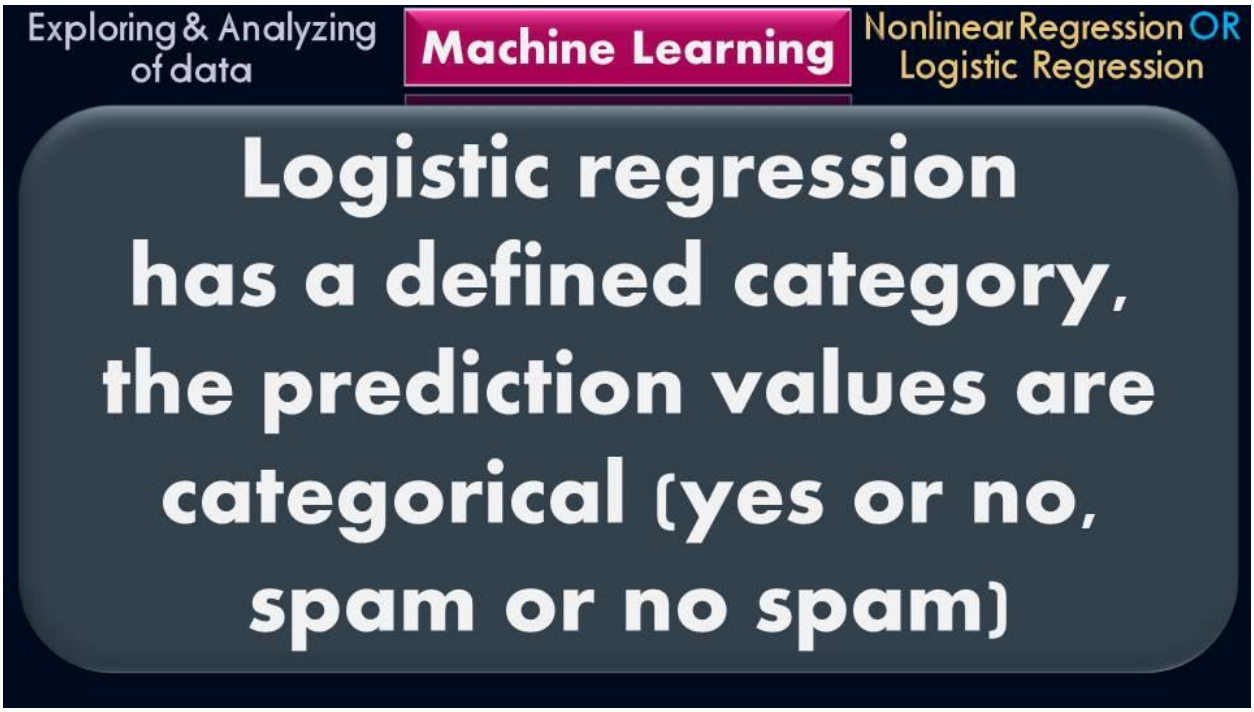

Exploring & Analyzing of data
Machine Learning
Linear Regression
b): Multiple linear regression
$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n$
a_i is coefficient of regression
x_i is independent variable
y is a dependent variable

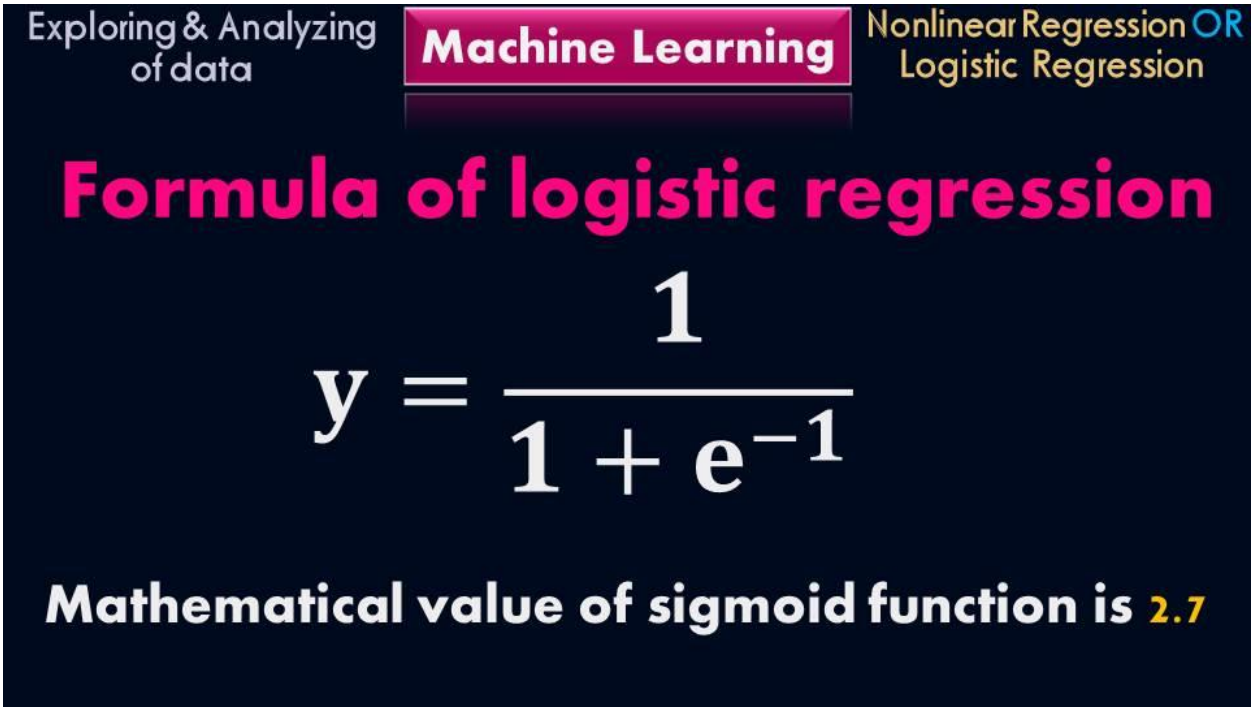

Exploring & Analyzing of data
Machine Learning
Linear Regression
$y = 0.5 + x_1 + 3x_2 + 2.5x_3 + \cdots + a_nx_n$
Where $a_2 > a_3$
Note: The importance of a data set is based on its weight (regression coefficient). In the above equation a_2(3) has more importance than a_3(2.5). Because 3>2.5

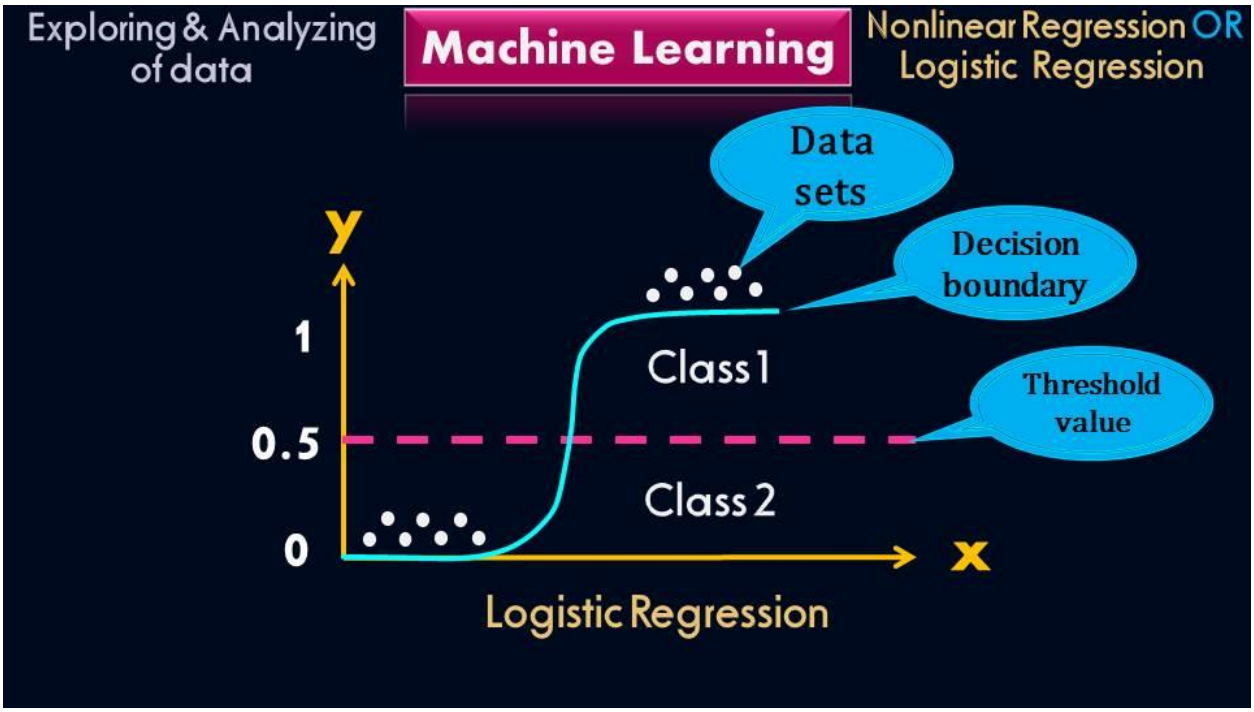

Exploring & Analyzing of data
Machine Learning
Linear Regression
Decision boundary
Data sets
y
x
Linear Regression

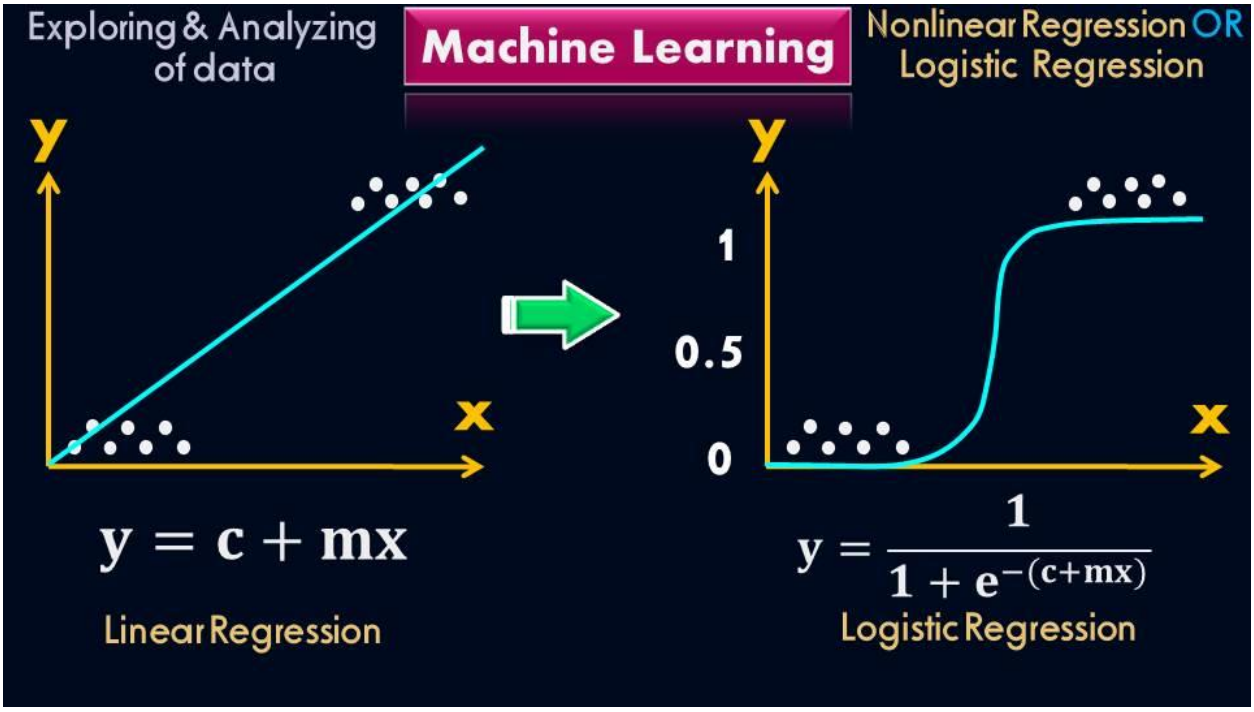

Exploring & Analyzing of data
Machine Learning
Nonlinear Regression OR Logistic Regression
Non linear regression or logistic regression is also called a sigmoid function having range from 0 to 1 with a threshold value 0.5


Exploring & Analyzing of data
Machine Learning
Linear Regression
Linear regression has no defined category, it's used to do prediction based on continues values. The output can be any number. Its disadvantages are; that this model is prone to over fitting and is useless when there's a non linear relationship between the dependent and independent variables.


Exploring & Analyzing of data
Machine Learning
Nonlinear Regression OR Logistic Regression
Logistic regression has a defined category, the prediction values are categorical (yes or no, spam or no spam)


Exploring & Analyzing of data
Machine Learning
Nonlinear Regression OR Logistic Regression
Formula of logistic regression
$y = \frac{1}{1 + e^{-1}}$
Mathematical value of sigmoid function is 2.7


Exploring & Analyzing of data
Machine Learning
Nonlinear Regression OR Logistic Regression
Data sets
Decision boundary
Threshold value
y
1
0.5
0
Class 1
Class 2
x
Logistic Regression


Exploring & Analyzing of data
Machine Learning
Nonlinear Regression OR Logistic Regression
y
x
1
0.5
0
$y = c + mx$
Linear Regression
$y = \frac{1}{1 + e^{-(c+mx)}}$
Logistic Regression


Exploring & Analyzing of data
Machine Learning
Nonlinear Regression OR Logistic Regression
Applications of logistic regression:
Fraud detection (spam, no spam)
Diagnosis of diseases
Alert detection
Detection of plagiarism
Weather forecasting
Prediction of tomorrow's market