Module 2

COMMON Regression ALGORITHMS

Content of the Module

- Linear Models for Regression
  - Simple Linear Regression
  - Multiple Linear Regression

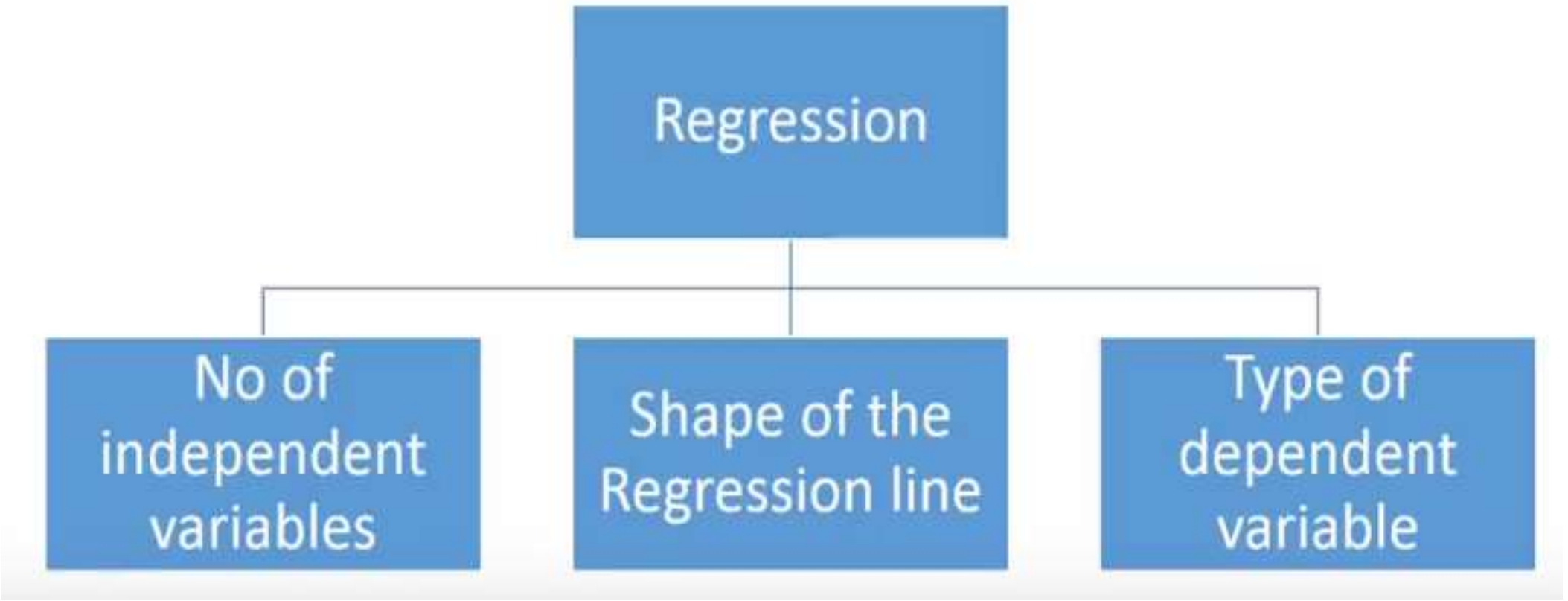
Regression

- **Regression analysis falls under supervised machine learning.**
- The system tries to predict a value for an dependent variable based on independent variable.
- **Characteristics of regression:**
  - The responses obtained from the model are always quantitative in nature.
  - The model can be constructed only if past data is available.
  - Basic tool in the area of machine learning used for predictions.
  - Developed in the field of statistics and ML borrowed it.

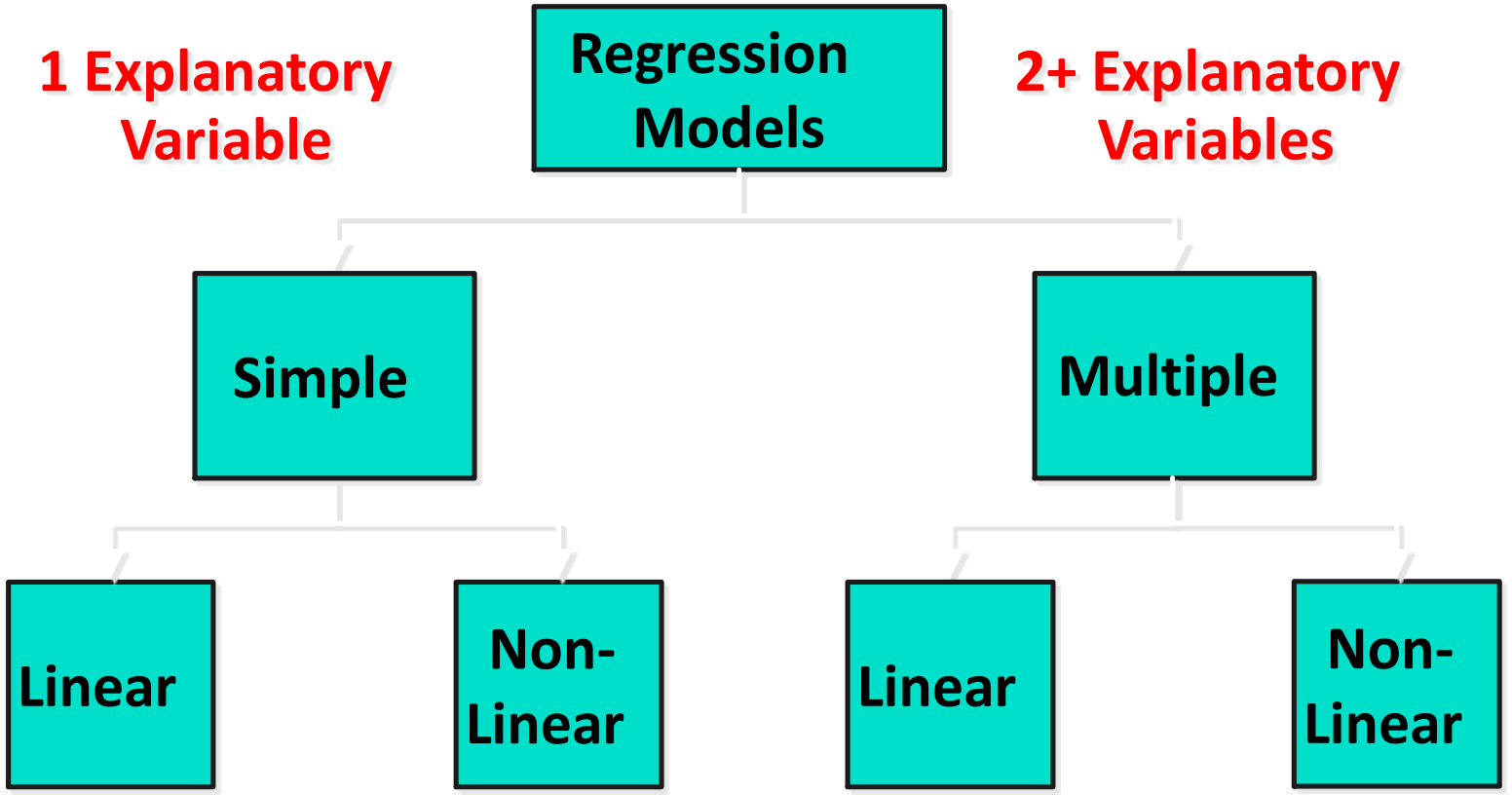
Regression

- **Basic Concept:**
- You fit a function on the available data and try to predict the outcome for the future or hold out data point.
- **Fitting function 2 purposes**
    - Interpolation – Missing data
    - Extrapolation – Future data

Type of Regression



Types of Regression Models



Linear Regression

- - Oldest
- - Simple
- - Widely used supervised ML algorithm
-

Linear Regression for predictive analysis.

Independent Variable & Dependent Variable

- area of house→ price of the house
- No. of hours of study → grade
- Sugar level→ diabetic patient
- Cholesterol Range -> heart disease

Linear Regression

- A linear regression line has an equation of the form

$$y = mx+b$$

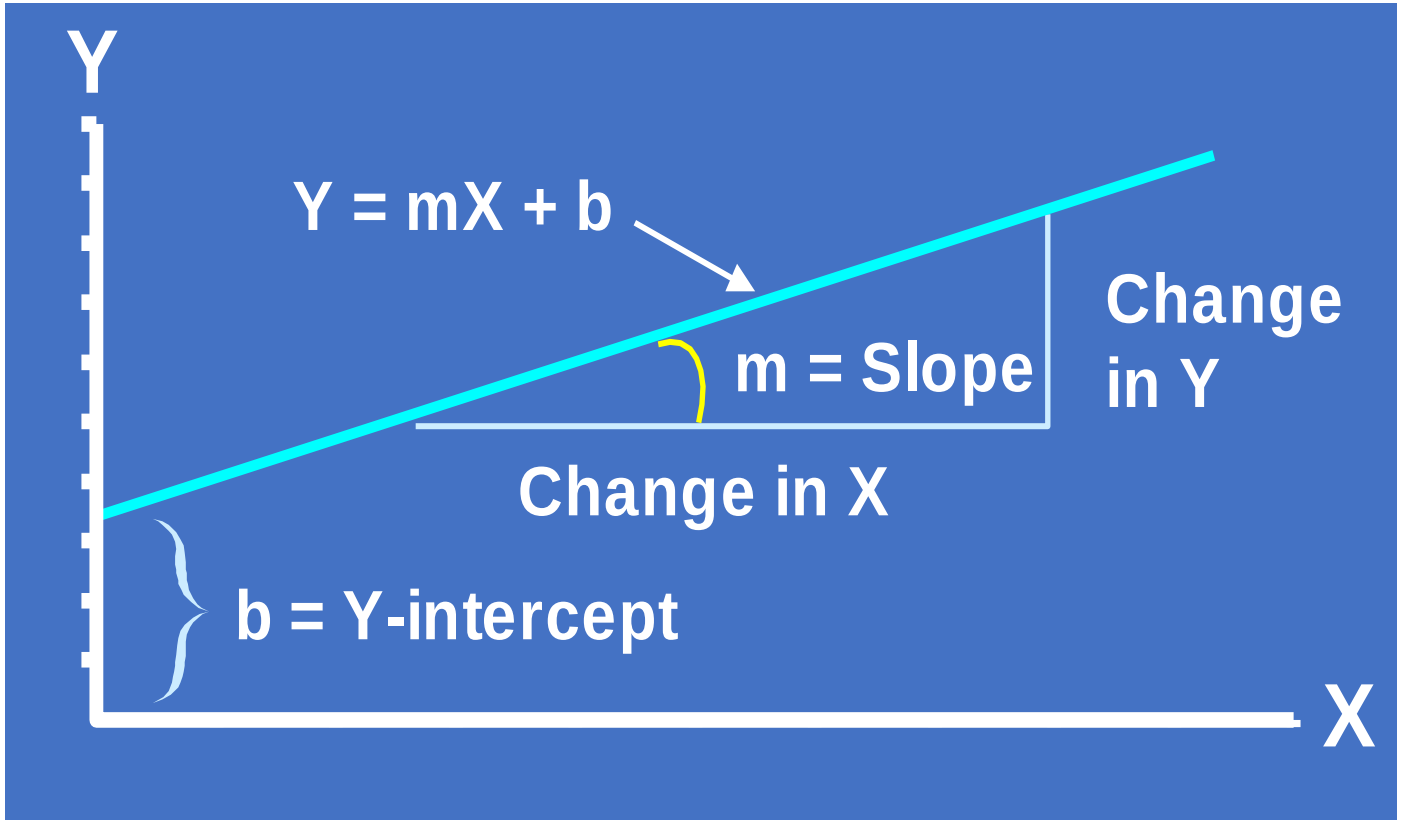
where

x-- is the explanatory variable (independent variable)

Y-- is the dependent variable,

m-- the slope of the line,

b ---is the intercept



Problem1:

- The weight of the person is related to his height. Find the relationship between height and weight using linear regression and also predict the weight of the person whose height is 170cm

Given Data

Height	Weight
151	63
174	81
138	56
186	91
128	47
136	57
179	76
163	72
152	62
131	48

Given Data ----Calculate

Height	Weight	Xi- Xmean	Yi-Ymean	(Xi- Xmean)(Yi-Ymean)	$(Xi-Xmean)^2$
151	63				
174	81				
138	56				
186	91				
128	47				
136	57				
179	76				
163	72				
152	62				
131	48				
Xmean= 153.8	Ymean= 65.3				

Given Data

Height	Weight	Xi- Xmean	Yi-Ymean	(Xi- Xmean)(Yi-Ymean)	$(Xi-Xmean)^2$
151	63	-2.8	-2.3	6.44	7.84
174	81	20.2	15.7	317.14	408.04
138	56	-15.8	-9.3	146.94	249.64
186	91	32.2	25.7	827.54	1036.8
128	47	-25.8	-18.3	472.14	665.64
136	57	-17.8	-8.3	147.14	316.84
179	76	25.2	10.7	269.64	635.04
163	72	9.2	6.7	61.64	84.64
152	62	-1.8	-3.3	5.94	3.24
131	48	-22.8	-17.3	394.44	519.84
Xmean= 153.8	Ymean= 65.3			∑ =2649.6	∑=3927.6

Calculate Slope value and Intercept value fo below formula

- **Y=b+mX**
- **Y=bo+b1x**
- **b(intercept)=**

$$b_0 = \frac{1}{n}\left(\sum Y_i - b_1 \sum X_i\right) = \bar{Y} - b_1 \bar{X}$$

m(Slope) =

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

Calculate Y …when X =170

- Given Formula
- $y=b+mx$
-

Calculate Y ...when X =170

- Given Formula
- y=b+mx

- y = -38.4551 + 0.6746 *170
- =76.23
-

Problem2:

- The rent of a property is related to its area. Given the area in square feet and rent in rupees. Find the relationship between area and rent using the concept of linear regression. Also predict the rent for a property of 790sq.ft.

Multiple Linear Regression

- Multiple Linear Regression is an extension of Simple Linear regression as it takes more than one predictor variable to predict the response variable. We can define it as:
- *Multiple Linear Regression is one of the important regression algorithms which models the linear relationship between a single dependent continuous variable and more than one independent variable.*
- **Example:**
- Prediction of CO_2 emission based on engine size and number of cylinders in a car.

steps to perform multiple linear Regression

- The steps to perform multiple linear Regression are almost similar to that of simple linear Regression.
- The Difference Lies in the evaluation.
- We can use it to find out which factor has the highest impact on the predicted output and how different variables relate to each other.

Multiple Regression

- *Here :* ***Y = b0 + b1 * x1 + b2 * x2 + b3 * x3 + …… bn * xn***

- **Y= Output/Response variable**
- **b_0, b_1, b_2, b_3 , b_n....= Coefficients of the model.**
- **x_1, x_2, x_3, x_4,...= Multiple Independent/feature variable**

Multiple Regression

- *Here :* ***Y = b0 + b1 * x1 + b2 * x2 + b3 * x3 + …… bn * xn***

- **Y= Output/Response variable**
- **b_0, b_1, b_2, b_3 , b_n....= Coefficients of the model.**
- **x_1, x_2, x_3, x_4,...= Multiple Independent/feature variable**

Steps Involved in any Multiple Linear Regression Model

- **Step #1:** Data Pre Processing
- Importing The Libraries.
- Importing the Data Set.
- Encoding the Categorical Data.
- Avoiding the Dummy Variable Trap.
- Splitting the Data set into Training Set and Test Set.
- **Step #2:** Fitting Multiple Linear Regression to the Training set
  Step #3: Predict the Test set results.