3. Regression.

Krushna Nagare

Last Minute Notes.

• **Introduction:**

Q- What is Linear regression? & application of LR.

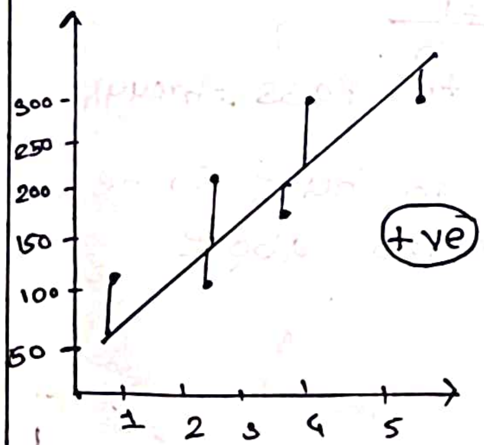
→ Liniear regression illustrate the relationship between two Variables or factors.

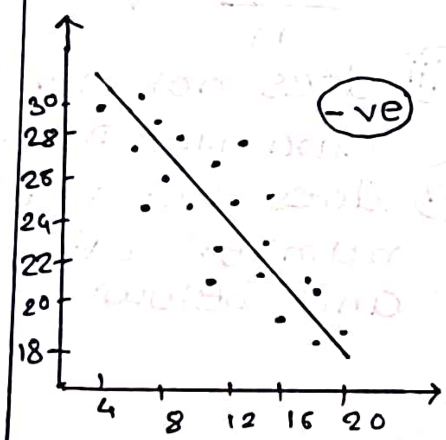
Y → dependent Variable

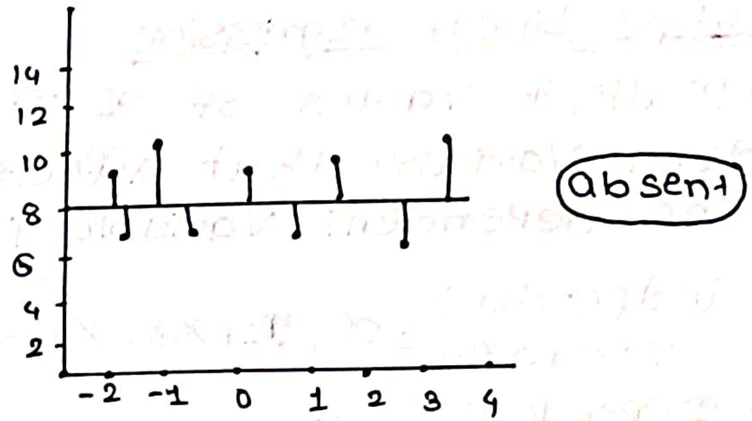
x → Independent Variable

$\boxed{y = b_0 + b_1 * x} \Rightarrow \boxed{y = c + mx}$

• Positive relationship
• negative relation ship.
• absent relationship







• Models – (i) Univariate Linear regression.
(ii) Multiple Linear regression.

• Univariate Linear regression:

▸ only one dependent & indepedent Variable

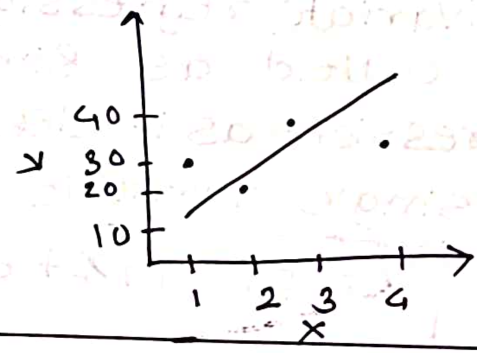
▸ Linear regression where only one (uni) input variable i.e. (x) known as univariate linear regression

$\therefore \boxed{y = \theta_1 + \theta_2 x}$

θ_1 - intercept

θ_2 = Coefficient of x (i.e. slope).

x	y
1	15
2	30
3	40
4	30



- **Multiple Linear regression**
  - multiple values of x i.e. independent variable that affects only one dependent variable i.e. (y).
  - independent variables = $x_1, x_2, x_3, x_4 \dots x_n$
  - Dependent variables = y.

$\therefore \boxed{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \dots \beta x_n}$

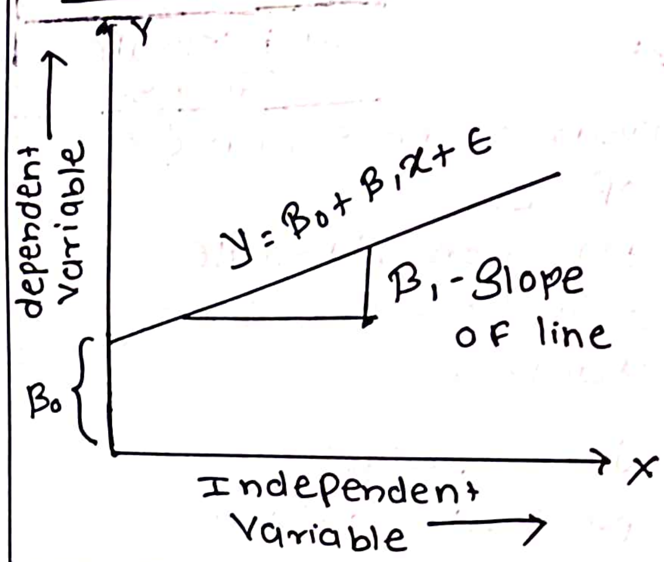
Examples:

1. Selling price of house depends upon multiple variables like location, no. of bedrooms, year of buiding
2. Height of child depend on harmons, nutrition, enviromental factors.

- **Least - square Method:**
  - Univariate regression can be also called as simple linear regression: as it depends only of single variable

eg:- $\boxed{Y = \beta_0 + \beta_1 x + \epsilon}$

ϵ - random error



- **Best fit regression**
  1. Regression through minimum sum of error.
  2. it must pass through centroid of data. $(\bar{X}, \bar{Y})$

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

  3. does not need to pass through maximum point
  4. does not need to have same number of points above and below.

$$\boxed{y = \beta_0 + \beta_1 x + \varepsilon}$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{x}$$

$$\beta_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

The least square method is the process of finding the best-fitted curve or line of best fit for a set of data Points by reducing the sum of the squares of offsets of the Points from the curve.

least square ⇒ Sum of all square of error is minimum

Cost Function:

MSE (Mean Squared error)

avarage of sum of squared of diffrence between actual Value and Predicted value Known as Mean squared error.

$$\boxed{MSE = \frac{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}{n}}$$

$$\boxed{MSE = \frac{SSE}{N}}$$

ME Mean error

Avarage of diffrance of actual Values and Predicted values

$$\boxed{ME = \frac{\sum_{i=1}^{n} (y_i - \bar{y}_i)}{n}}$$

MAE

The Avarage Absolute diffrance betn actual & predicted Values.

$$\boxed{MAE = \frac{\sum_{i=1}^{n} |y_i - \bar{y}_i|}{n}}$$

R-Squared (R2-Score)

Statastical measure that represent the goodness of fit of regression model

$$\boxed{R^2 = 1 - \frac{SSE}{Var(y)}}$$

NMSE (Normalized mean squared error)

$$NMSE = \frac{SSE}{Var(y)} = \frac{1}{Var(y)} \times \sum_{i=1}^{n} (y_i - \bar{y}_i)^2$$

RMSE (Root mean squared error)

$$\boxed{RMSE = \sqrt{MSE}}$$

$$\boxed{SSE = \sum_{i=1}^{n} (y_i - \bar{y}_i)^2}$$

• Problems

- **Performance evaluation:**

(i) Various ML algorithm are evaluated using Performance metrics.

(ii) After training Performance is done on Testing dataset.

(iii) confusion matrix used to evaluate model.

★ **Confusion Matrix**

Actual

Predicted	1	0
1	TP	FP
0	FN	TN

(i) Accuracy :-

$$A = \frac{TP+TN}{Total}$$

A

P		
	TP	FP
	FN	TN

(ii) precision

$$P = \frac{TP}{TP+FP}$$

TP	FP
FN	TN

(iii) Recall

$$R = \frac{TP}{TP+FN}$$

TP	FP
FN	TN

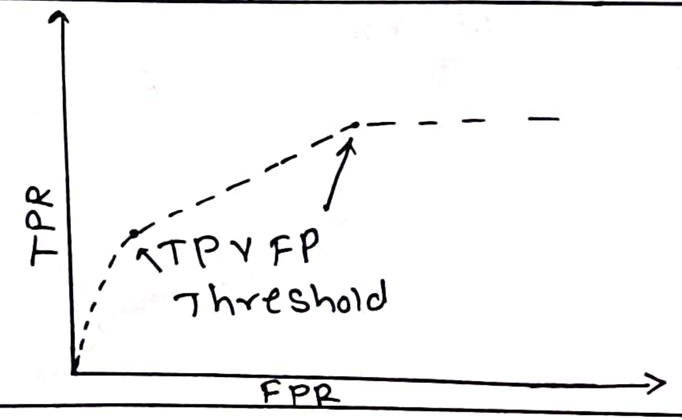
(iv) F1-Score

$$F1 = \frac{2 \times Precision \times recall}{(Precision + Recall)}$$

- **Roc curve**

Reciver oprating charecterstic curve is a graphical Plot that explain the diagnostic ability of a binary classifier system.

TPR (True +ve rate) $= \frac{TP}{TP+FN}$	FPR (false +ve rate) $= \frac{FP}{FP+TN}$



• Problems

• Optimization of Simple linear regression with Gradient descent - Example

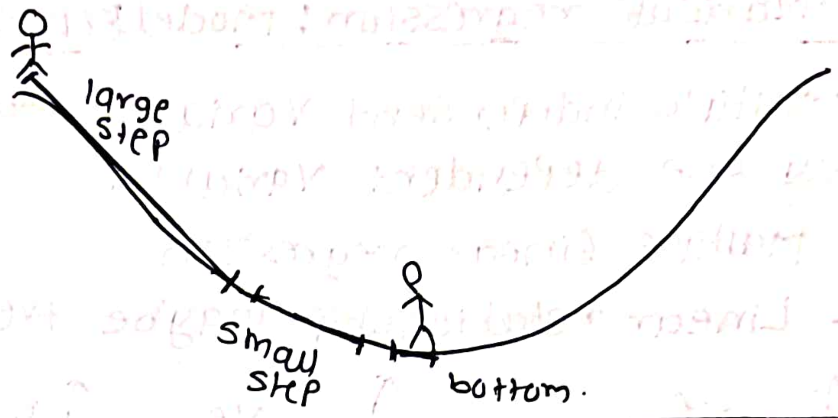
• **Gradient Descent Algorithm**

it is an iterative optimization algorithm to find minimum of fun^n. Here that function is Loss fuction

• By using gradient descent we perform derivative of Variable to get the best-fit

• Model makes prediction → error is calculater → Model updated → error is reduced → next prediction → prediction

• Example:- Person wan'ts to reach bottom of velly with no sense of direction. when the slope (steep) is inclide he takes bigger stees & when the value of D (slope) is low he takes the smaller steps.



Steps

① $m=0 \quad c=0 \quad L=0.0001$

if L is high the low accurcy (Person will keep oscilating)

② $Dm = \frac{1}{n}\sum_{i=0}^{n} 2(y_i - (mx_i + c))(-x_i)$

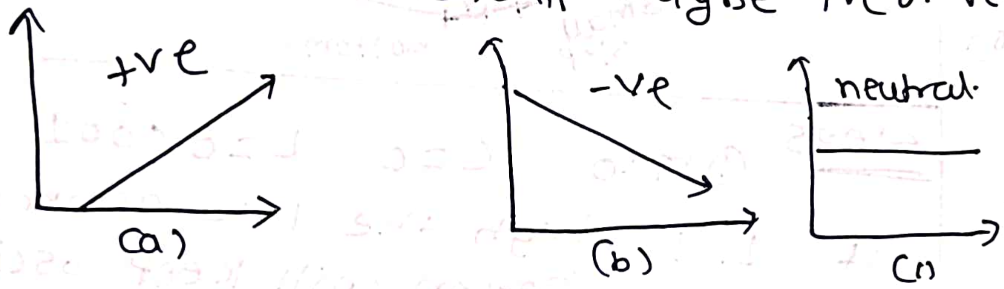
$Dm = \frac{1}{n}\sum_{i=0}^{n} 2(y_i - \bar{y})(-x_i)$

$Dc = \frac{-2}{n}\sum_{i=0}^{n} (y_i - \bar{y})$

③ Repeat function till L is very low $\rightarrow 0$

...variable regression: model representation

- Multiple independent variable for only one dependent variable.
- Multiple linear regression
- Linear relationship maybe +ve or -ve



- representation

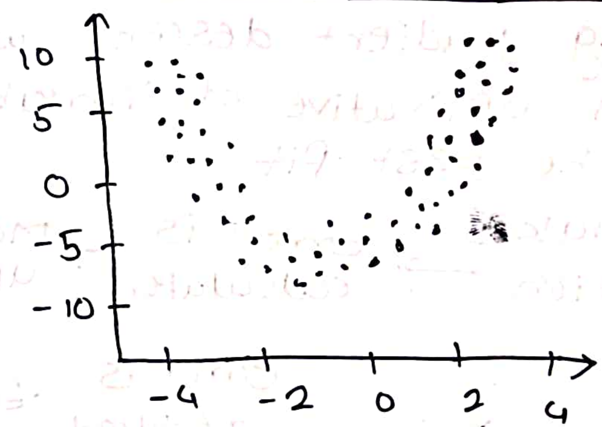
$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \dots b_nx_n$$

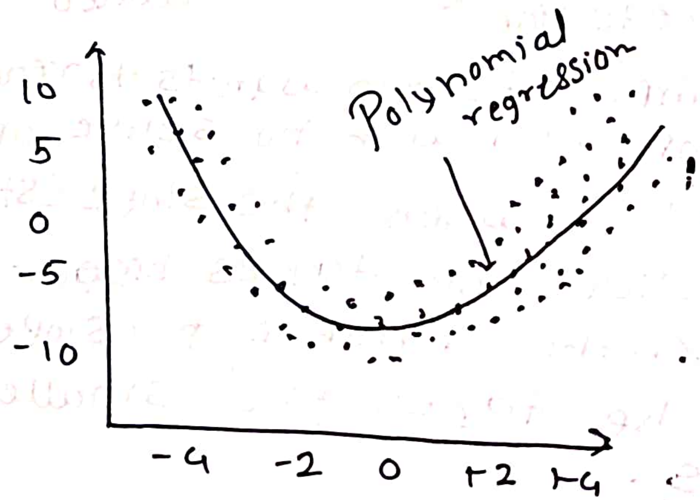
- Multiple Linear regression attempt to model the relationship between two or more features and a response by fitting a linear equation to observed data.

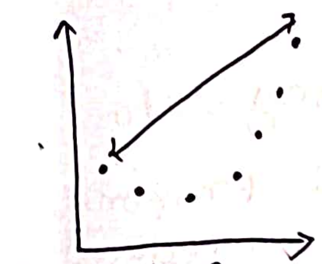
Predictive ability	Interpredictive.
Test error is combination of two or more variable	coefficient is assigned to every element/ variable.
accuracy can be ↑ by losing some amount of bias	

Polynomial Regression-

- relation is kind of regression analysis in which the relationship between independent variable x & dependent y is in n^{th} degree of polynomial.







(a) SLR

(b) PR

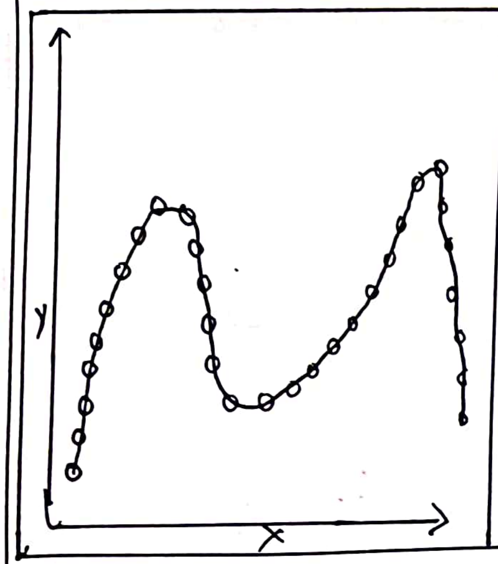
- High Accuracy
- For non-linear dataset
- low loss fuction
- Less error
- Special case of mutiple Linear regression.

$$Y = b_0 + b_1 x^2 + b_2 x^3 + b_3 x^4$$

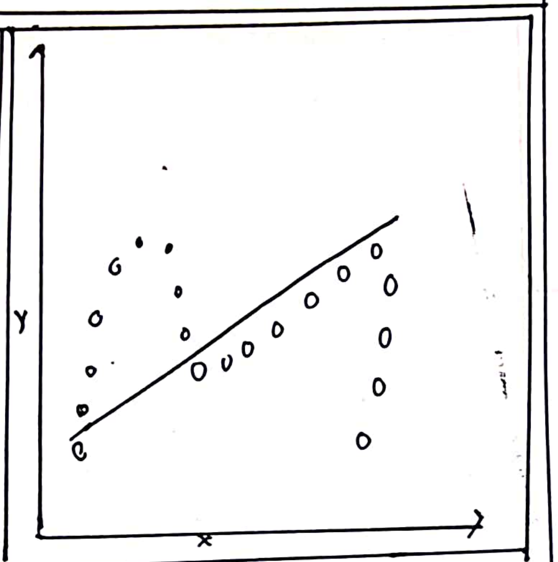
Advantages	Disadvantages
• Best approximation • Widerange of funⁿ • wide range of curve	• Even one or two outliner can affect result • Very sensitive • fewer validation tools.

Q- write Short Note on Sigmoid.

• Underfitting & overfitting

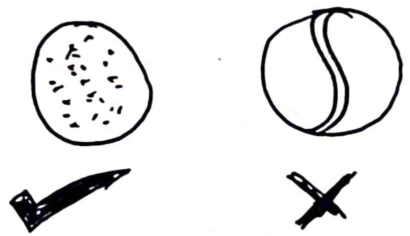


overfitting



underfitting

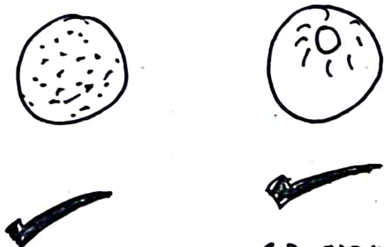
① Ball
→ Sphare
→ raidus - 5cm
→ color - red
→ matrial - Plastic



Plastic ball

tennis ball

② Ball
→ sphare



plastic ball

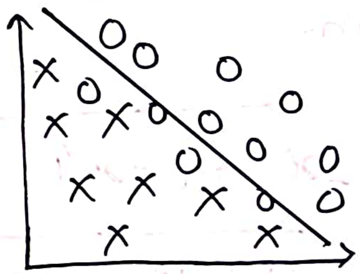
orange

overfitting

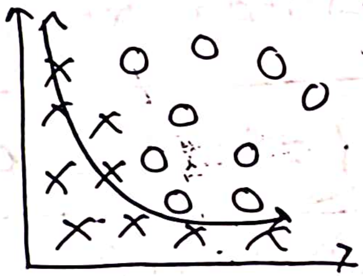
- Trained with so many data
- Cannot categorized propery
- use of linear algo to reduce overfitting.

underfitting

- Trained with very low data
- does not fit well
- more feature selection to ovecome underfitting



underfitting



best-fit

overfit

Bias v/s Variance

Let consider we have two values

Predicted ← → actual

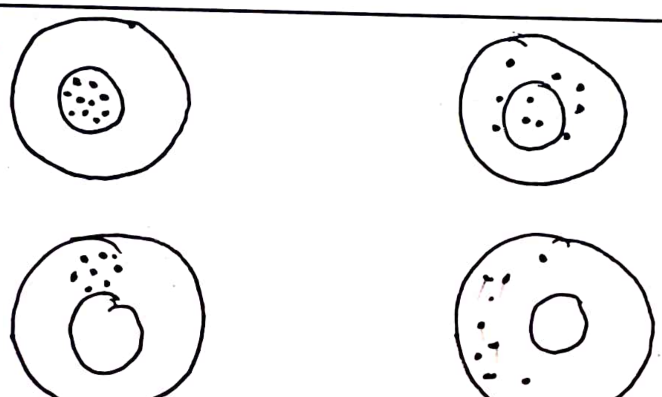
Bias = gap between two values

high Bias → Large gap	Low Bias → smaller gap

• Variance refers to the spread of expected values in relation to one another.

low variance → All predicted values we will see in a group

High Variance → Predicted values will be scattred.



Bias	Variance
(i) Diffn between Prediction & actual Values	spread of our data is caued as varince of model
(ii) Model with high bias pay littel attesion to training data set	High variance model pay close attension to training data.
(iii) It leads to high error on training & test data.	Perform very well on traing dataset