PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

Semester : VI Subject :CSC601 Data Analytics and Visualization Academic Year: 2023- 2024

Module-2

Regression Models

2.1 Introduction to simple Linear Regression

Basic algorithm for Data Science & Deep learning

1) What problems are we solving
2) Geometric Intuition
3) Mathematical Intuition.

Machine Learning
- Supervised
  - ~~Regr~~ Regression
    - output features with continuous values
    - e.g. Linear Regression
  - Classification
- Unsupervised

Types of Data
- Qualitative ~~Quantitative~~
  - Nominal — e.g. gender, eye color
  - Ordinal — Feedback in scale 1 to 10 - Education level (Primary, Secondary, Higher)
- Quantitative
  - Discrete
  - Continuous

Discrete - ~~Height~~ ~~of~~ Total no of students present in class
- no of employees in a company.

Continuous - Height of a person, age, weight
- Time taken to finish a task

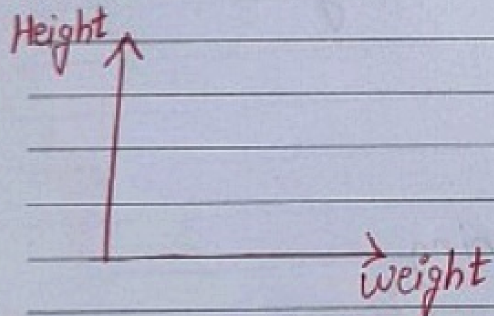
PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

Semester : VI Subject :CSC601 Data Analytics and Visualization Academic Year: 2023- 2024

What is Linear Regression

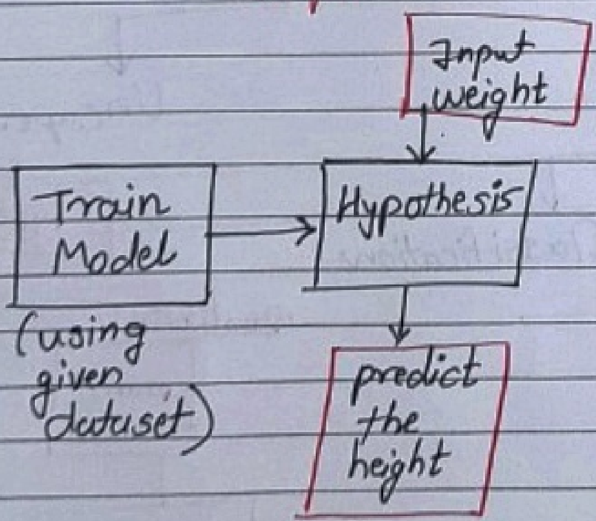
- This mathematical intution helps in understanding the working of the model designed. Hyper parameter tunning can also be done with the help of linear regression.

- Problem Statement of linear regression



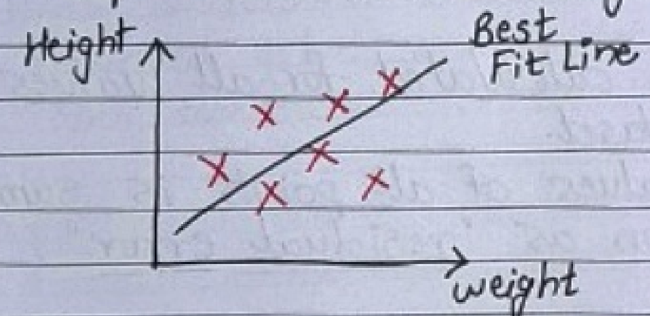
Features

Weight (kg) ↑ Independent Feature	Height (cm) ↑ Dependent Feature
73	150
63	155
78	170
82	165

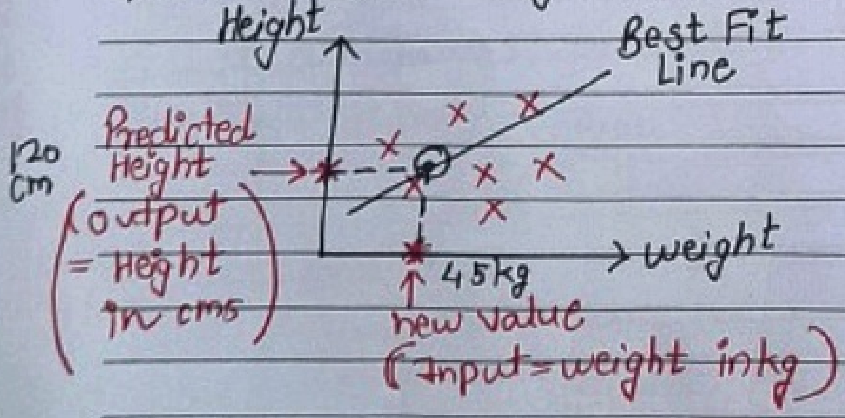


The output is a continuous value which height of a person.

Let's plot the values of given dataset in the graph.



Based on the values from given dataset linear regression draws best fit line. Once the best fit ~~to~~ line is drawns we can pass new value ~~to~~ the model which is input weight & predict the height with the help of "best fit line".



How to draw best fit line?

- To find the best fit line for the given dataset first difference between actual value and predicted value needs to be calculated.
- This difference is known as "residual error".

Residual Error = Actual ~~Error~~ Value − Predicted Value

- There is possibility to have multiple lines and from those lines we need to select the best fit line.
- Criteria for selection of the best line is to select the line where sum of all residual errors of all points present in the dataset in minimum.

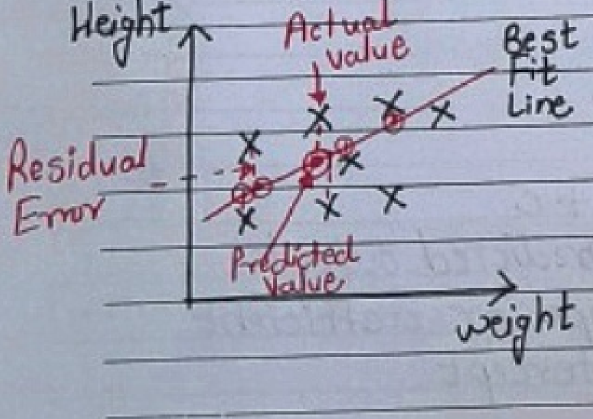


Figure shows difference actual value & predicted value as per best fit line.
The difference between these two is residual error. for that point.

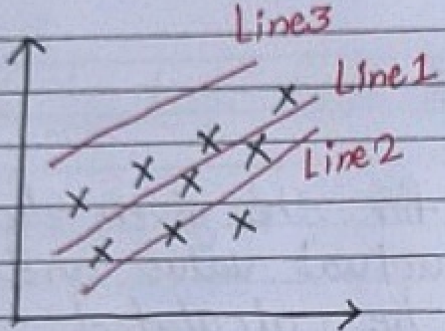


Figure shows line1, line2 and line3 are the lines available for selection of best fit line where the line with minimum residual error is selected.

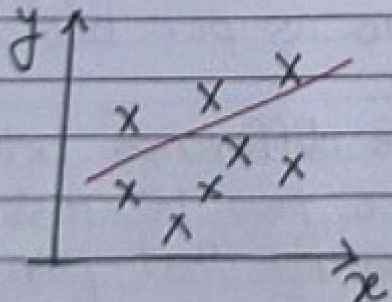
If the residual error value is zero it is called "overfitting condition." Normally the residual error never reaches to exact zero.

The Problem Statement of Linear Regression

"To find the best fit line with minimal residual error"

Solution

Equation of best fit line



$\hat{y} = mx + c$

$\hat{y}$ = predicted output

m = slope or coefficient

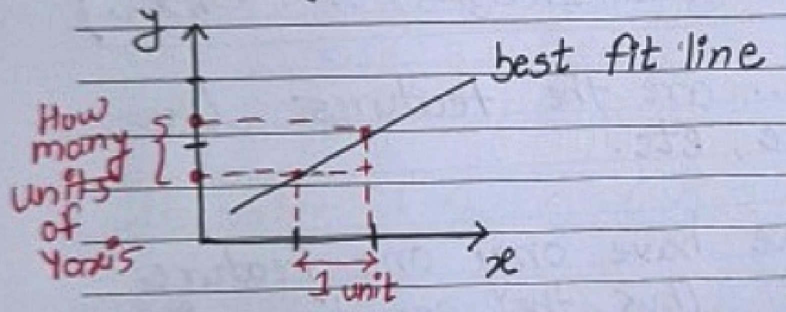
c = intercept

PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

Semester : VI Subject :CSC601 Data Analytics and Visualization Academic Year: 2023- 2024

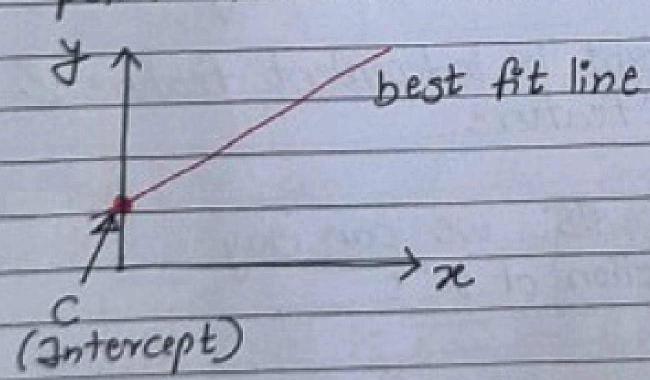
What is slope or coefficient?

If we increase the value of x axis by 1 unit then we need to calculate the slope to identify how many units are increased in y.



What is intercept?

When the value of x is zero where or at what point the best fit line intercept the y axis.



We need to write hypothesis for the given problem statement

The equation of hypothesis by Andrew NG

$$\boxed{h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots \theta_n x_n}$$

here $x_1, x_2, x_3 \dots$ are the features like height, weight, age, etc.

For our problem we have only one feature which is weight. Thus, the equation of hypothesis is as follows.

$$\boxed{h_\theta(x) = \underbrace{\theta_0}_{\text{Intercept}} + \overbrace{\theta_1}^{\text{coefficient}} x_1}$$

This will be hypothesis created by "Best fit Line"

In our scenario weight is independent feature & height is dependent feature.

For the above hypothesis we can say {"y is a linear function of x"}

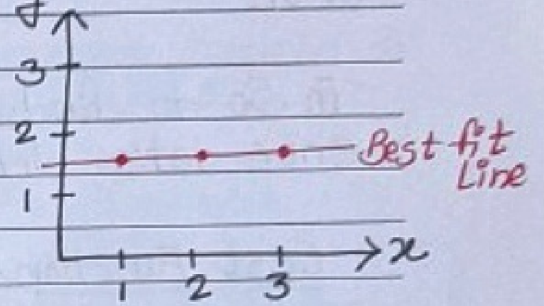
PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

Semester : VI Subject :CSC601 Data Analytics and Visualization Academic Year: 2023- 2024

Example,

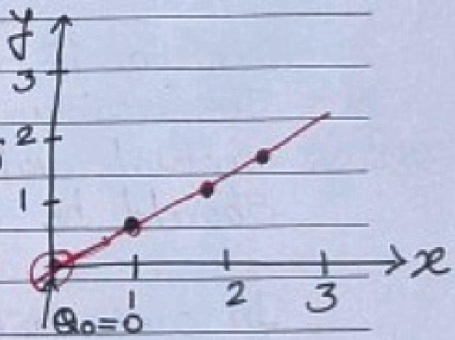
$$\boxed{h_\theta(x) = \underbrace{\theta_0}_{\text{Intercept}} + \underbrace{\theta_1}_{\text{coefficient}} x_1}$$

If value of $\theta_0 = 1.5$ & $\theta_1 = 0$



$x=1$, $h_\theta(x) = 1.5 + 0 = 1.5$
$x=2$, $h_\theta(x) = 1.5 + 0 = 1.5$
$x=3$, $h_\theta(x) = 1.5 + 0 = 1.5$

If value of $\theta_0 = 0$ & $\theta_1 = 0.5$



~~x=~~ $x=1$, $h_\theta(x) = 0 + 0.5 \times 1 = 0.5$
$x=2$, $h_\theta(x) = 0 + 0.5 \times 2 = 1$
$x=3$, $h_\theta(x) = 0 + 0.5 \times 3 = 1.5$

If $\theta_0 = 0$ then the Best fit line passes through point 0 where $x=0$ & $y=0$.

Now the aim of the problem statement is to find best fit line with minimum residual error.

So the point is how to find this best fit line or how to determine it's position.

We can not go on changing the values of θ_0, θ_1 & try to plot the lines then calculating residual error.

So to find best fit line with minimum residual error let's define a cost function

Cost Function

Our aim is to minimise the residual error
The residual error is difference between actual value & predicted value. This difference should be as low as possible.

In our case, predicted point we get from $h_\theta(x)^i$ for all data points in the dataset.

And y^i is the actual value
Next we need to sum all the residual error.
After summing up all the values we take average of the total so we divide it by m.

m in our case is total number of points.

PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

Semester : VI Subject :CSC601 Data Analytics and Visualization Academic Year: 2023- 2024

$$\boxed{\sum_{i=1}^{m} \frac{1}{2m} \left(h_\theta(x)^i - y^{(i)}\right)^2}$$

The answer for $(h_\theta(x)^i - y^{(i)})$ can be positive or negative. So the ensure to get answer of it as a positive integer we perform square of the answer i.e. $(h_\theta(x)^i - y^{(i)})^2$.

To calculate average of all residual errors in the line we perform 1/m. But instead of 1/m we use 1/2m. ~~when~~ While calculating the slope, it is required to calculate the ~~slope~~ differentiation

In differenciation calculation of x^2 with x.

$$\frac{d(x^2)}{dx} = 2x^{2-1} = 2 \cdot x$$

So the reduce the complexity of the operation we divide the equation by 1/2.

$$\boxed{J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x)^i - y^i\right)^2}$$ Squared Error Function

To reduce the value of cost function as low as possible is our aim.

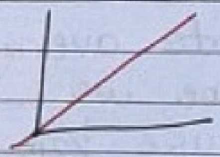
Hypothesis ~~Cost Function~~

$h_\theta(x) = \theta_0 + \theta_1 x_1$

Cost Function

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x)^i - y^i)^2$$

Let's assume $\theta_0 = 0$



Line is passing through intersept

So hypothesis will be

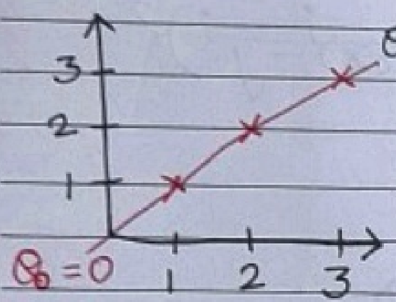
$h_\theta(x) = \theta_1 x_1$

1) We need to find how the value of θ_1 will change to find the best fit value
2) Try to find how hypothesis and cost function are related.

Now continue with hypothesis

For $\theta_0 = 0$, $h_\theta(x) = \theta_1 x_1$

Now we will see how to change value of θ_1



Data Points Given
(1,1)
(2,2)
(3,3)

$\boxed{\text{when } \theta_1 = 1}$

$h_\theta(x) = \theta_1 \cdot x_1$

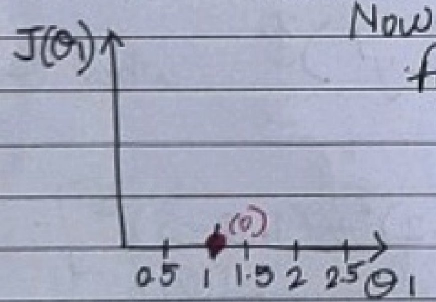
$h_\theta(1) = 1 \cdot 1 = 1$

$h_\theta(2) = 1 \cdot 2 = 2$

$h_\theta(3) = 1 \cdot 3 = 3$

To find correct value of θ_1 is important to find correct best fit line.

Now let's find the value of cost function $J(\theta_1)$



$\theta_1 = 1$

$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x)^i - y^i)^2 \quad \text{when } \theta_1 = 1$$

$$= \frac{1}{2 \cdot 3} \cancel{\Sigma} \left[(0)^2 + (0)^2 + (0)^2 \right] = 0$$

when $\theta_1 = 1$

we have 3 data points given

$\theta_1 = \overset{0.5}{\cancel{1}}$, $J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h\theta(x)^i - y^i)^2$

$= \frac{1}{2m} [($

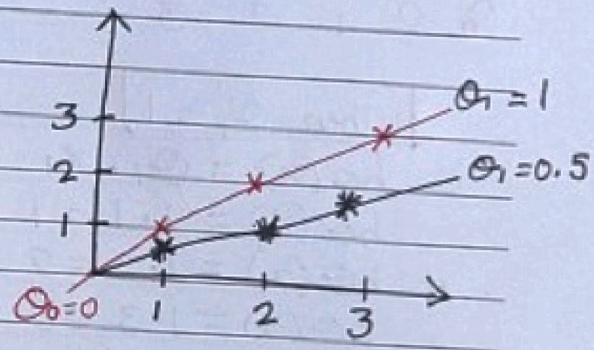
$\boxed{\text{When } \theta_1 = 0.5}$

$h\theta(x) = \theta_1 \cdot x_1$

$h\theta(1) = 0.5 \times 1 = 0.5$

$h\theta(2) = 0.5 \times 2 = 1.0$

$h\theta(3) = 0.5 \times 3 = 1.5$

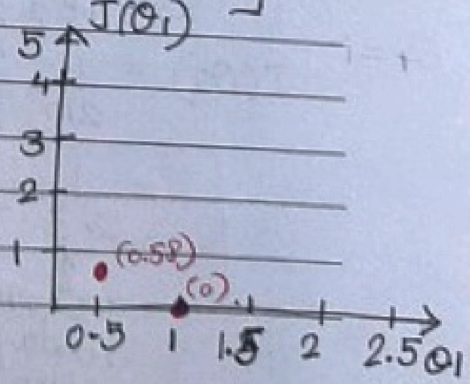


Now calculate cast funⁿ $J(\theta_1)$ for $\theta_1 = 0.5$

$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h\theta(x)^i - y^i)^2$ wher $\theta_1 = 0.5$

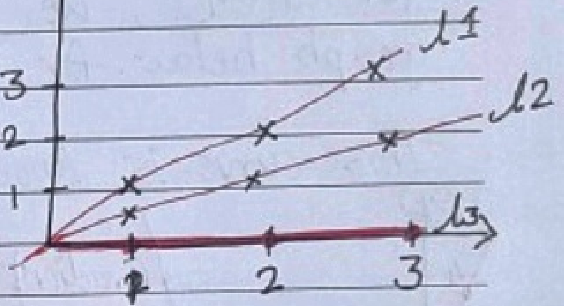
$= \frac{1}{6} [(0.5-1)^2 + (1-2)^2 + (1.5-3)^2]$

$= \frac{1}{6} [3.5] \approx 0.58$



$\boxed{\text{When } \theta_1 = 0.0}$

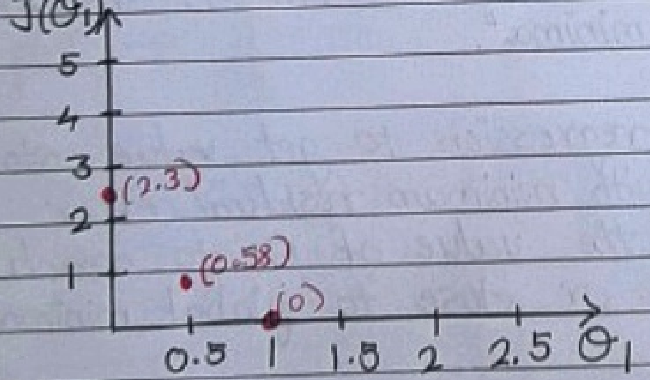
$$h\theta(x) = \theta_1 \cdot x_1$$
$$h\theta(1) = 0 \times 1 = 0$$
$$h\theta(2) = 0 \times 2 = 0$$
$$h\theta(3) = 0 \times 3 = 0$$



$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h\theta(x)^i - y^i)^2 \quad \text{where } \theta_1 = 0.0$$

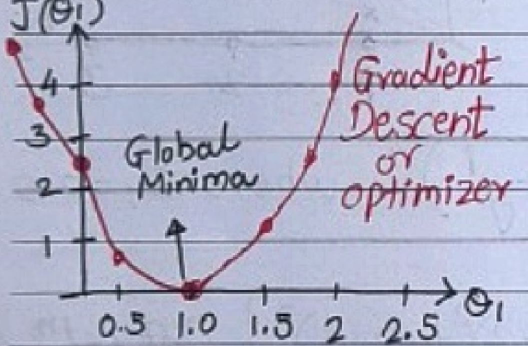
$$= \frac{1}{6}\left[(0-1)^2 + (0-2)^2 + (0-3)^2\right] \quad \text{Diff bet}^n \ l1 \ \& \ l3$$

$$= \frac{14}{6} \approx 2.3$$



Likewise if we go on calculating the values for $J(\theta_1)$, we get a curve as shown in graph below. for different values of θ_1

This curve is known as "Gradient Descent" or optimizer



From the Gradient Desent select point where difference between actual point & predicted point least.
Such point is called "Global Minima"

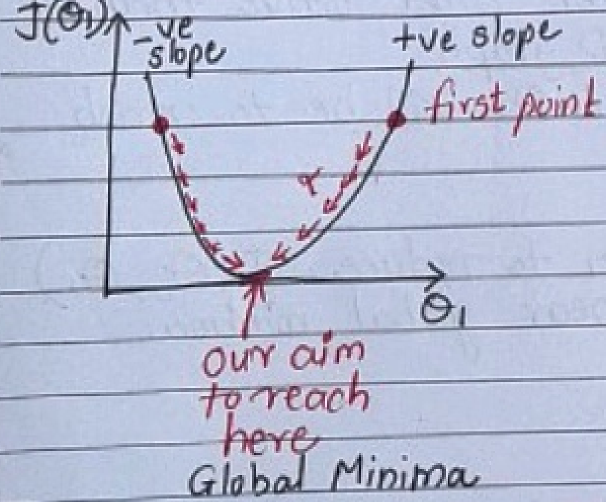
we need to train our model till it reaches close to "global minima".

For for linear regression to get ~~value~~ ~~eto~~ best fit line with minimum residual error we need to change the value of θ_1 to reach to global minima. or close to global minima.

PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

Semester : VI Subject :CSC601 Data Analytics and Visualization Academic Year: 2023- 2024

How to identify if we have reach close to global minima & how the value of θ_1 will be updated ?



To change the value θ_1 till it reaches to global minima, repeat convergence theorem is used.

1) +ve slope → if x increases y also increases

2) -ve slope → if x increases y decreases

3) For updating θ_1
 - find derivative or slope

4) α - learning rate decides the speed with which value of θ_1 will change.

5) $\frac{\partial}{\partial \theta_j}(J(\theta_1))$ - derivative of cost fun^n which is slope

6) α - small value, slow changes in value of θ_1

Repeat Convergence (
{

$$\theta_j = \theta_j - \alpha \underbrace{\frac{\partial}{\partial \theta_j}(J(\theta_1))}_{\text{slope}}$$

}

5 - 4 = 1 (small value) reduced value

+ve slope $\theta_1 = \theta_1 - \alpha(+ve)$

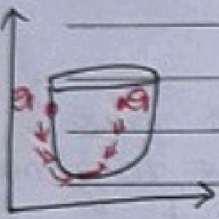
5 - (-4) = 5 + 4 = 9

-ve slope $\theta_1 = \theta_1 - \alpha(-ve)$

PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

Semester : VI   Subject :CSC601 Data Analytics and Visualization   Academic Year: 2023- 2024

1) Start with some θ_0 & θ_1



In our case we have considered $\theta_0 = 0$
If it is any other other value than 0
we will have 3D graph.
But aim of θ_0 & θ_1 will be to reach global minima.

2) Keep changing θ_0, θ_1 to reduce $J(\theta_0, \theta_1)$ until we reach near global minima

3) Convergence Theorem

{

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} (J(\theta_0, \theta_1))$$

for $J = 0$ & 1

}