Data Science and Artificial Intelligence

Machine Learning

Regression

Lecture No. 04

By- SIDDHARTH SABHARWAL SIR

Recap of Previous Lecture

Class | Notes | DPP

Topic

B Formulae

Topic

Representation X, Y

Topic

2D data → attributes features.

Topic

Topic

Topics to be Covered

Topic Gradient descent

Topic Loss fxn in matrix

Topic R^2 : coeff of determination

Topic

Topic

About the Faculty

- AIR 1 GATE 2021, 2023 (ECE).
- AIR 3 ESE 2015 ECE.
- M.Tech from IIT Delhi in VLSI.
- Published 2 papers in field of AI-ML.
- Paper 1 : Feature Selection through Minimization of the VC dimension.
- Paper 2 : Learning a hyperplane regressor through a tight bound on the VC dimension.

By- SIDDHARTH SABHARWAL SIR

Don't Limit Your Challenges
Challenge Your Limit

@Lionhearted

REVISION

How the data is represented in matrix format

D dimension data

Data Point	D_1	D_2	-- D_D	y
1	x_1^1	x_1^2	x_1^D	y_1
2	x_2^1	x_2^2	x_2^D	y_2
3	⋮	⋮	⋮	⋮
4				
⋮				

$$X = \begin{bmatrix} 1 & x_1^1 & x_1^2 & \cdots & x_1^D \\ 1 & x_2^1 & x_2^2 & \cdots & x_2^D \\ 1 \\ \vdots \\ \vdots \end{bmatrix}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_D \end{bmatrix}$$

The final expression of Beta matrix..

$$\beta = (X^T X)^{-1} (X^T Y) \Rightarrow \text{imp}$$

Regression.

REVISION

Regression and Classification...

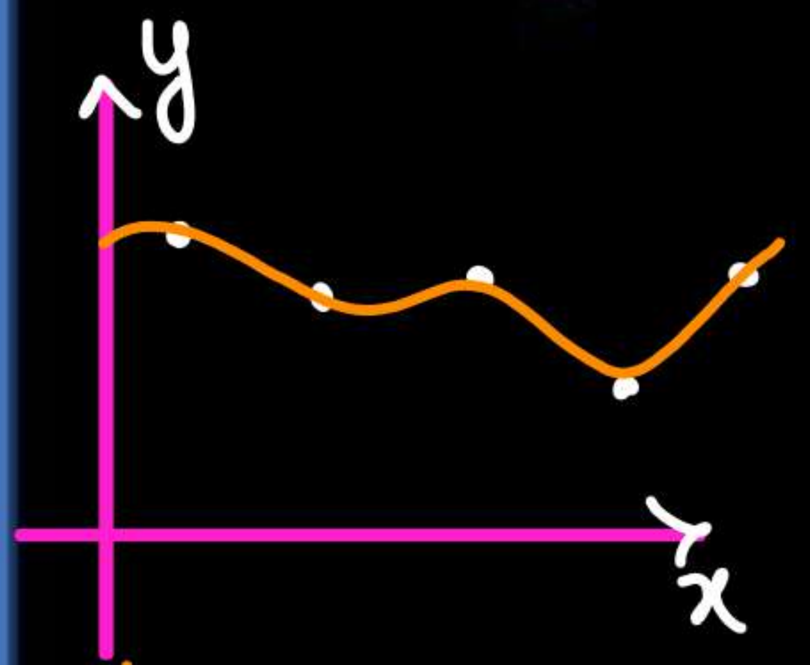
2 Type of data

Humidity	Temp	Wind speed	Amt of Rain
40	30	15	25
30	40	20	33
10	15	20	42
.	.	.	41
.	.	.	46

→ label Y

⇓

- Numbers
- Real values



- Here we want a fxn Relating Y and x

Classification

REVISION

Regression and Classification...

2 Type of data

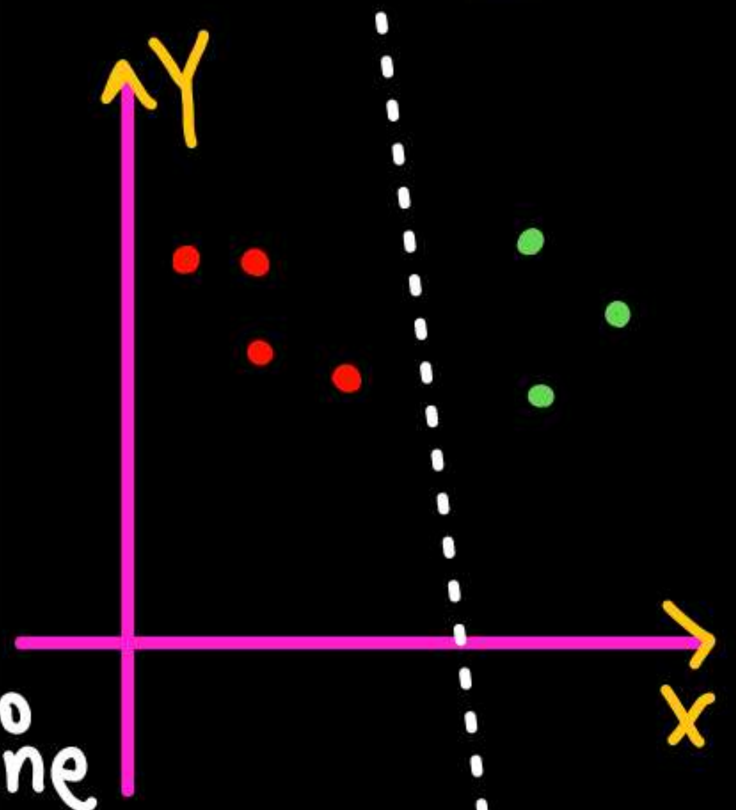
Humidity	Temp	Wind speed	Rain
40	30	15	Yes
30	40	20	No
10	15	20	Yes
'	'	'	Yes
'	'	'	No

Rain → label

Y: Yes/No

Categorical Nature

Here we want a line that seperates.



Red-green Point

Regression :- Since we are finding a linear relation b/w y and x

$$y = \beta_0 + \beta_1 x^1 + \beta_2 x^2 + \cdots$$

So this is linear Regression

What is the primary goal of linear regression in the context of 1D data?

find linear relation b/w y and x, and

A) To classify data into categories

B) To predict a continuous output variable based on an input feature

C) To reduce the dimensionality of the data

D) To cluster data points into groups

In simple linear regression (1D), the relationship between the independent variable x and the dependent variable y is modeled as:

A) $y = mx + c + \epsilon$

B) $y = mx^2 + c$

C) $y = \sin(x) + c$

D) $\log$ form: $y = \log(x) + c$

data

x	y
x_1	y_1
x_2	y_2
x_3	y_3
$\vdots$	$\vdots$

Relation by y and x

$\Rightarrow y = mx + c \rightarrow$ Predicted value fxn

$$y = mx + c + \varepsilon$$

$\varepsilon \rightarrow$ (error term, which is very small)

What does the term $\boxed{\text{"residual"}}$ represent in linear regression?

A) The slope of the regression line

$\boxed{\text{B) The difference between the predicted value and the actual value}}$

C) The intercept of the regression line

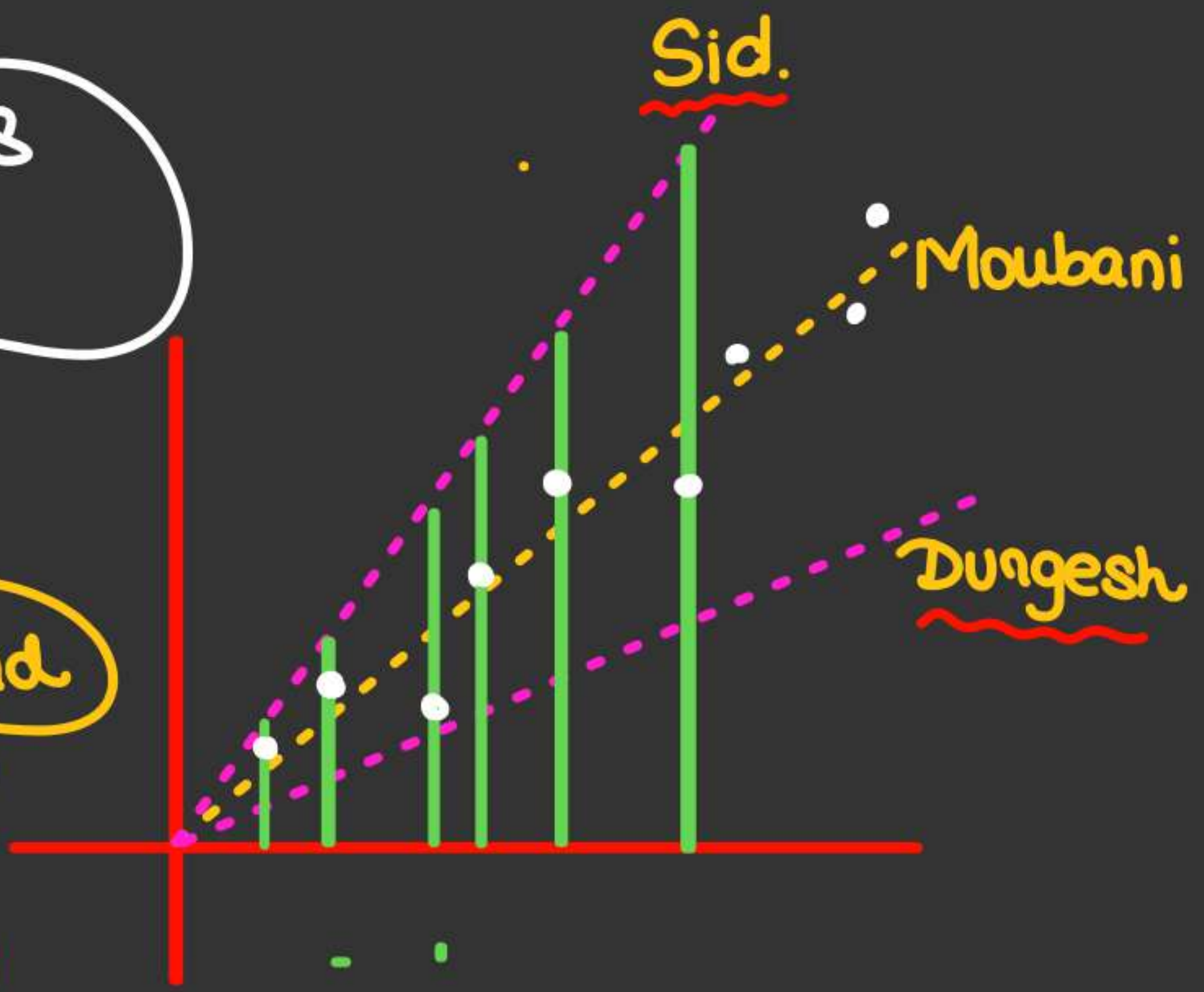
D) The correlation coefficient

Best line : data Ke beech मे cross Kare, Residual ±ve.

Q.

Which is best :-

Yellow line : Best understand data Pattern



Residual : (actual - Predicted)

Sid : Residual always -ve

Dungesh : Residual always +ve

Moubani : Residual ±ve

Skip

In the context of linear regression, what does R^2 (R-squared) measure?

A) The slope of the regression line

B) The proportion of variance in the dependent variable explained by the independent variable

C) The residual error

D) The intercept of the regression line

What happens to the regression line if all the data points lie exactly on a straight line?

A) The residuals will be zero

B) The slope will be zero

C) The intercept will be zero

D) The R^2 value will be zero

Skip

Which of the following is a limitation of linear regression?

A) It cannot handle categorical variables

B) It assumes a linear relationship between variables

C) It is sensitive to outliers

D) All of the above

Given the following data points for x and y:

x	y
1	2
2	3
3	5
4	6

H.W

We min sum of Residual^2.

$$B = (X^T X)^{-1} (X^T Y)$$

$$\text{Loss fxn} = \sum_{i=1}^{N} (y_i - y_i')^2$$

we use this eq

$$\frac{\partial L}{\partial m} = 0, \frac{\partial L}{\partial c} = 0 \text{ eq form}$$

Find the slope (m) and intercept (c) of the best-fit line $y = mx + c$ using the least squares method.

For a linear regression model, the following statistics are given:

H.W

- Mean of x ($\bar{x}$) = 5
- Mean of y ($\bar{y}$) = 10
- Variance of x (σ_x^2) = 4
- Covariance between x and y ($\mathrm{Cov}(x, y)$) = 6

Find the slope (m) and intercept (c) of the regression line.

For a linear regression model, the sum of squared residuals (SSR) is 50, and the total sum of squares (TSS) is 200. What is the value of R^2?

Skip

A dataset has the following properties:

1D data with x, y as data values.

- Number of data points $(n) = 10$
- $\sum x = 30, \sum y = 50$
- $\sum xy = 200, \sum x^2 = 120$

Find the slope (m) of the regression line.

1D data

$$[X^T X]\beta = (X^T Y)$$

$$\rightarrow \begin{bmatrix} \sum 1 & \sum x \\ \sum x & \sum x^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$

$$\begin{bmatrix} 10 & 30 \\ 30 & 120 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 50 \\ 200 \end{bmatrix}$$

$$[X^T X]\beta = (X^T Y)$$

1D data

$$\begin{bmatrix} \sum 1 & \sum x \\ \sum x & \sum x^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$

$$\begin{bmatrix} 10 & 30 \\ 30 & 120 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 50 \\ 200 \end{bmatrix}$$

- $\beta_0 = c = 0$
- $\beta_1 = m = 5/3$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \left(\frac{1}{120 \times 10 - 30 \times 30}\right) \begin{bmatrix} 120 & -30 \\ -30 & 10 \end{bmatrix} \begin{bmatrix} 50 \\ 200 \end{bmatrix} \Rightarrow \frac{1}{300} \begin{bmatrix} 0 \\ 500 \end{bmatrix}$$

$\beta_0 = 0$, $\beta_1 = 5/3$

1D data

$$\begin{matrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \\ x_4 & \vdots \\ \vdots & \vdots \\ x_{10} & y_{10} \end{matrix}$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_{10} \end{bmatrix}$$

$$X^T = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_{10} \end{bmatrix}$$

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_{10} \end{pmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_{10} \end{bmatrix}$$

$$\Rightarrow \begin{Bmatrix} \sum_{i=1}^{10} 1 & \sum_{i=1}^{10} x_i \\ \sum_{i=1}^{10} x_i & \sum_{i=1}^{10} x_i^2 \end{Bmatrix}$$

How to represent the Loss function in the matrix format

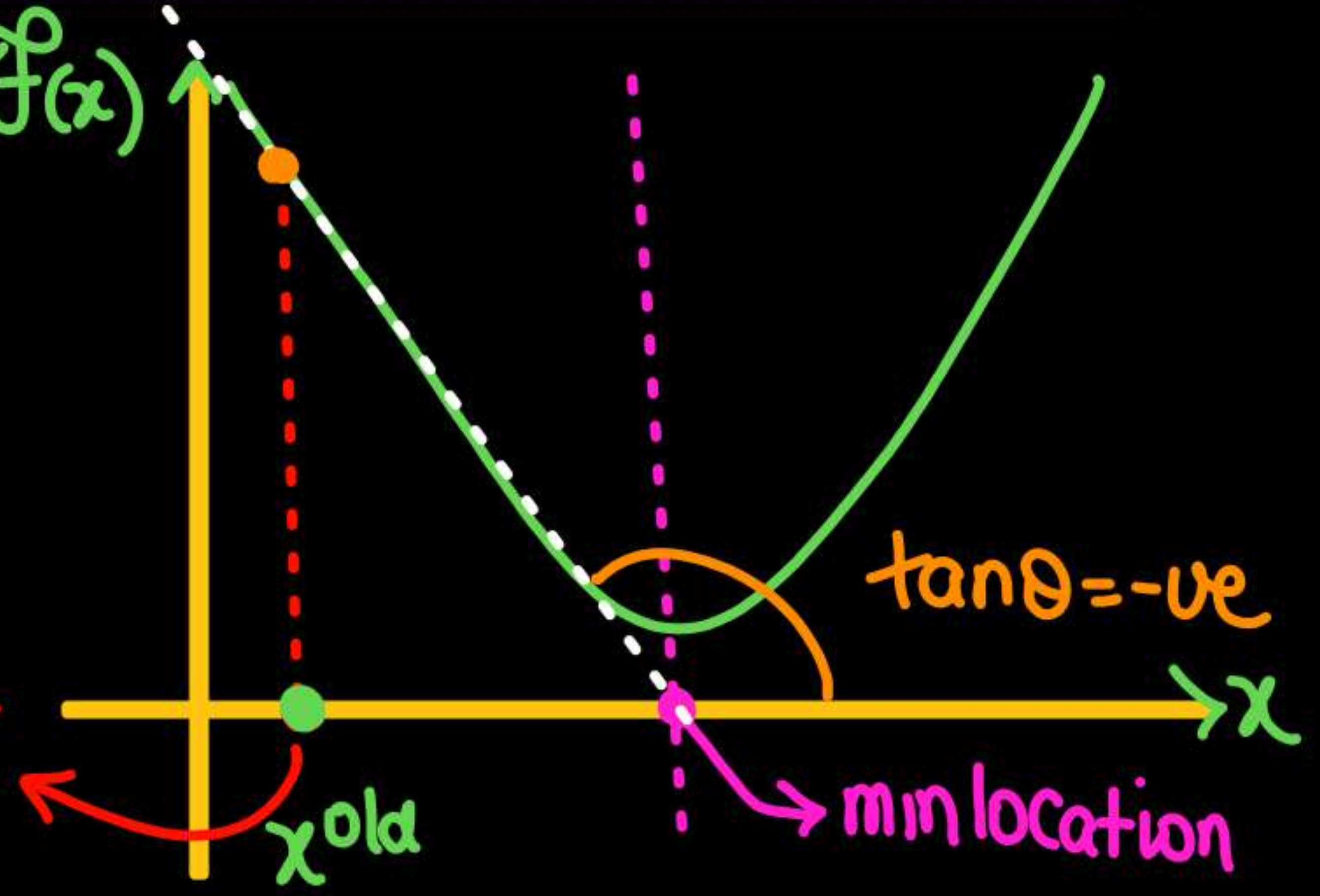
Gradient Descent

⇓

Maths

We start to find min location:

1) we start @ any Random x location



Const

$$\rightarrow x^{next} = x^{old} - \eta \left(\left. \frac{df(x)}{dx} \right|_{x^{old}} \right)$$

How to represent the Loss function in the matrix format

Gradient Descent

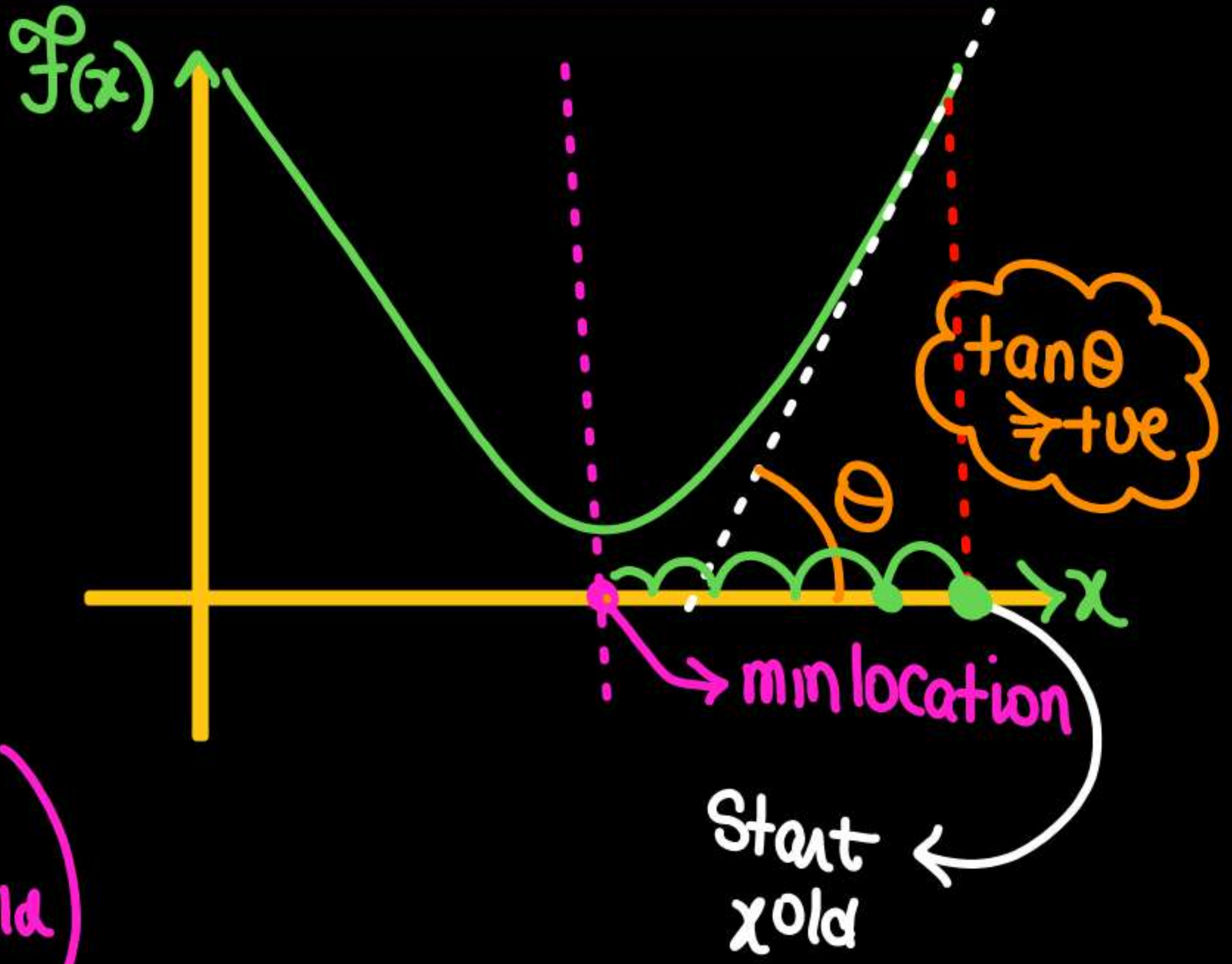
⇓

Maths

We start to find min location:

1) we start @ any Random x location

$$\rightarrow x^{next} = x^{old} - \underset{\text{Const}}{\eta} \left(\frac{df(x)}{dx}\Big|_{x^{old}} \right)$$



GAATE 2026 FODENGE

How to represent the Loss function in the matrix format

Gradient descent

↳ maths Concept to find min of a fxn.

we have a fxn $f(x)$

Step 1: we start with any Random x value

Step 2: Iteration 1: $x^{old} = \text{Value}$

$$x^{new} = x^{old} - \eta \left(\frac{\partial F(x)}{\partial x} \bigg|_{x^{old}} \right)$$

learning Rate

How to represent the Loss function in the matrix format

Step 3 Iteration 2

Here iteration 1 Result x^{old}

$$\Rightarrow \left(x^{new} = x^{old} - \eta \left. \frac{\partial F(x)}{\partial x} \right|_{x^{old}} \right)$$

So on

@min algorithm stops: @min $df(x)/dx = 0$

How to represent the derivative of L by Beta in matrix format

Happy Mode

How to represent the derivative of L by Beta in matrix format

1D data

x	y
x_1	y_1
x_2	y_2
x_3	y_3
x_4	y_4

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

$$y = \beta_0 + \beta_1 x$$

$$L = \sum_{i=1}^{4} (y_i - y_i')^2$$

$$L = \sum_{i=1}^{4} (y_i - (\beta_0 + \beta_1 x_i))^2$$

$$\frac{\partial L}{\partial \beta} = \begin{bmatrix} \partial L/\partial \beta_0 \\ \partial L/\partial \beta_1 \end{bmatrix} \Rightarrow \begin{bmatrix} -2\left(\sum_{i=1}^{4} y_i - \beta_0 - \beta_1 x_i\right) \\ -2\left(\sum y_i x_i - \beta_0\right. \end{bmatrix}$$

$$\frac{\partial L}{\partial \beta} = \begin{bmatrix} \partial L/\partial \beta_0 \\ \partial L/\partial \beta_1 \end{bmatrix} \Rightarrow \begin{bmatrix} -2\left(\sum_{i=1}^{4} y_i - \beta_0 - \beta_1 x_i\right) \\ -2\left(\sum_{i=1}^{4} y_i x_i - \beta_0 x_i - \beta_1 x_i^2\right) \end{bmatrix}$$

$$\frac{\partial L}{\partial \beta} = -2 \begin{bmatrix} \sum_{i=1}^{4} y_i - \beta_0 \sum_{i=1}^{4} 1 - \beta_1 \sum_{i=1}^{4} x_i \\ \sum_{i=1}^{4} x_i y_i - \beta_0 \sum_{i=1}^{4} x_i - \beta_1 \sum_{i=1}^{4} x_i^2 \end{bmatrix}$$

$$\frac{\partial L}{\partial \beta} = -2\begin{bmatrix} \sum_{i=1}^{4} y_i - \beta_0 \sum_{i=1}^{4} 1 - \beta_1 \sum_{i=1}^{4} x_i \\ \sum_{i=1}^{4} x_i y_i - \beta_0 \sum_{i=1}^{4} x_i - \beta_1 \sum_{i=1}^{4} x_i^2 \end{bmatrix}$$

$$= -2\left\{ \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} - \begin{bmatrix} \sum 1 & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \right\}$$

$$\frac{\partial L}{\partial \beta} = \begin{bmatrix} \partial L / \partial \beta_0 \\ \partial L / \partial \beta_1 \end{bmatrix} = -2\left\{ X^T Y - (X^T X)\beta \right\}$$

GAATE 2026 FODENGE

What is gradient descent method

Q $f(x) = 3x^2 + 4x + 6$

$x^{start} = 3$

$\eta = .1$

find x @ end of 1st and 2nd iteration

$\frac{df(x)}{dx} = 6x + 4$

It 1: $x^{old} = 3$

$$x^{new} = x^{old} - \eta \left.\frac{df(x)}{dx}\right|_{x^{old}}$$

$$= 3 - (.1)\left[6x+4\right]_{x=3}$$

$$= .8$$

It 2: $x^{old} = .8$

$$x^{new} = x^{old} - \eta \left.\frac{df(x)}{dx}\right|_{x^{old}}$$

$$= .8 - (.1)\left[6x+4\right]_{x=.8}$$

$$= -.08$$

Loss fxn

$$\Rightarrow \mathcal{L} = \left(15 + 20\beta_0^2 + 15\beta_1^2 + 30\beta_0\beta_1\right)$$ → Randomly Question

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

Let we start $(\beta_0 = \beta_1 = 1)$, $\eta = .1$, find $\begin{matrix}\beta_0 \\ \beta_1\end{matrix}$ after 2 iteration

Sol.

$$\beta^{new} = \beta^{old} - \eta \left.\frac{\partial L}{\partial \beta}\right|_{\beta^{old}}$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}^{new} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}^{old} - \eta \begin{bmatrix} \partial L/\partial \beta_0 \\ \partial L/\partial \beta_1 \end{bmatrix}_{\beta^{old}}$$

$$\frac{\partial L}{\partial \beta_0} = 40\beta_0 + 30\beta_1$$

$$\partial L/\partial \beta_1 = 30\beta_1 + 30\beta_0$$

Sol. $\beta^{new} = \beta^{old} - \eta \left.\frac{\partial L}{\partial \beta}\right|_{\beta^{old}}$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}^{new} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}^{old} - \eta \begin{bmatrix} \partial L/\partial \beta_0 \\ \partial L/\partial \beta_1 \end{bmatrix}_{\beta^{old}}$$

$$\boxed{\begin{aligned} \frac{\partial L}{\partial \beta_0} &= 40\beta_0 + 30\beta_1 \\ \partial L/\partial \beta_1 &= 30\beta_1 + 30\beta_0 \end{aligned}}$$

It 1 $\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}^{old} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $\quad \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}^{new} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - .1 \begin{bmatrix} 70 \\ 60 \end{bmatrix} \Rightarrow \begin{bmatrix} -6 \\ -5 \end{bmatrix}$

It$_2$: $\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}^{old} = \begin{bmatrix} -6 \\ -5 \end{bmatrix}$, $\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}^{new} = \begin{bmatrix} -6 \\ -5 \end{bmatrix} - .1 \begin{bmatrix} -390 \\ -330 \end{bmatrix} \Rightarrow \begin{bmatrix} 33 \\ 28 \end{bmatrix}$

#Q. Let's consider regression in one dimension, so our inputs $x^{(i)}$ and outputs $y^{(i)}$ are in $\mathbb{R}$.

(a) (4 points) Linny uses regular linear regression. Given the following dataset, 1D data

(Home work)

$$D = \{((1), 1), ((2), 2), ((3), 4), ((3), 2)\}$$

What value of θ and θ_0 optimize the mean squared error of hypotheses of the form $h(x; \theta, \theta_0) = \cancel{\theta x + \theta_0 ?}$ $(\theta x + \theta_0)$

β_1

β_0

Linear Regression

$$y = \left(\beta_0 + \beta_1 x^1 + \beta_2 x^2 - - -\right)$$

* β_0 : intercept
  - Bias Term

Coeff of 1st dimen

" " 2nd "

⋮

#Q. Consider a one-dimensional regression problem with training data $\{x_i, y_i\}$. We seek to fit a linear model with ~~no bias term~~ Only Bias term.

Sol

data

x_1 y_1

x_2 y_2

$\vdots$ $\vdots$

x_N y_N

model $y = \beta_0$

→ we start from Begining

$$L = \sum_{i=1}^{N} (y_i - \beta_0)^2 \qquad \frac{\partial L}{\partial \beta_0} = 0, \quad \cancel{2} \sum_{i=1}^{N} (y_i - \beta_0)(\cancel{-1}) = 0$$

$$\Rightarrow \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} \beta_0 = 0$$

$$\Rightarrow \sum_{i=1}^{N} y_i - \beta_0 \sum_{i=1}^{N} 1 = 0$$

$$\left\{ \beta_0 = \frac{\sum_{i=1}^{N} y_i}{N} \right\}$$

#Q. Consider a one-dimensional regression problem with training data $\{x_i, y_i\}$. We seek to fit a linear model with no bias term:

Sol ⇒ data

$x_1 \quad y_1$

$x_2 \quad y_2$

$\vdots \quad \vdots$

$x_N \quad y_N$

model $y = \cancel{\beta_0} + \beta_1 x$

$y = \beta_1 x$

→ we start from Begining

$$L = \sum_{i=1}^{N} (y_i - \beta_1 x_i)^2$$

$$\min L : \frac{\partial L}{\partial \beta_1} = 0 \Rightarrow \cancel{2} \sum_{i=1}^{N} (y_i - \beta_1 x_i) x_i = 0$$

So. $$\sum_{i=1}^{N}\left(y_i x_i - \beta_1 x_i^2\right) = 0$$

$$\beta_1 \sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} y_i x_i$$

$$\Rightarrow \beta_1 = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}$$

#Q. Consider the following 4 training examples:

X	Y
-1	1
0	2
1	4
2	6

We want to learn a function $f(x) = ax + b$ which is parametrized by (a, b). Using squared error as the loss function, which of the following parameters would you use to model this function.

H.W

#Q. The linear regression model $y = a_0 + a_1x_1 + a_2x_2 + ... + a_px_p$ is to be fitted to a set of N training data points having p attributes each. Let X 1 point be $N \times (p + 1)$ vectors of input values (augmented by 1's), Y be $N \times 1$ vector of target values, and θ be $(p + 1) \times 1$ vector of parameter values $(a_0, a_1, a_2, ..., a_p)$. If the sum squared error is minimized for obtaining the optimal regression model, which of the following equation holds?

H.W.

(a) $X^TX = XY$

(b) $X\theta = X^TY$

(c) $X^TX\theta = Y$

(d) $X^TX\theta = X^TY$

Consider the function $J(w) = w_1^2 + w_2^2 - 6w_1 + 8w_2 - 9$. Answer questions (1-6):

[redacted]

H.W

gradient descent.

$$\Rightarrow \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}^{new} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}^{old} - \eta \begin{pmatrix} \partial J/\partial w_1 \\ \partial J/\partial w_2 \end{pmatrix} @ \text{ old } w_1, w_2$$

4) Start with the initial guess of $[w_1, w_2] = [5, 5]$. Take the value of learning rate = 0.3. The value of w_1 after 2 iterations of gradient descent will be ________.

THANK - YOU