Data Science and Artificial Intelligence

Machine Learning

Regression

Lecture No. 03

By- SIDDHARTH SABHARWAL SIR

Recap of Previous Lecture

Topic — Loss function → RSS → $\partial L/\partial m = 0$, $\partial L/\partial c = 0$

Topic — Single dimension Linear Regression

Topic — direct formula

Topic

Topic

Topics to be Covered

Topic — Question

Topic — Single dimension → multidimension

Topic — General : linear regression

Topic — Matrix.

Topic


Internal Motivation
PUSH YOURSELF,
BECAUSE NO ONE ELSE
IS GOING TO DO IT
FOR YOU.
SUCCESS.com

REVISION

1. What is the Loss Function

$$\mathcal{L} = RSS = \sum_{i=1}^{N}\Big(y_i - \underbrace{y_{\text{predicted}_i}}_{\Downarrow}\Big)^2$$

$y = mx + c$

Predicted value

$\Rightarrow \underbrace{mx_i + c}$

3. Direct formulae for M and C.

- $m = \dfrac{Cov(x, y)}{Var\, x}$ $\quad C = \bar{y} - m\bar{x}$

Mean $\Rightarrow \bar{x} \Rightarrow \dfrac{\text{Sum of all values}}{\text{Number of values}}$

4. Covariance :

$$\mathrm{Cov}(x,y) = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{N-1}$$

5. Variance : $\mathrm{Var}(x) = \sigma_x^2 = \dfrac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}$

Linear Regression

• data have single input in each datapoint

⇒ Thus data has single dimension

6. So the data that we were using has Single number of dimensions and the straight line obtained by liner regression has 2 [m and c] number of parameters.

$y = mx + c$

data

(x_1, y_1)

(x_2, y_2)

(x_3, y_3)

input

label/Output

- $$L = \sum_{i=1}^{N} \left(y_i - (m x_i + c)\right)^2$$

$$\frac{\partial L}{\partial m} \Rightarrow \sum_{i=1}^{N} x_i y_i - m \sum_{i=1}^{N} x_i^2 - c \sum_{i=1}^{N} x_i = 0$$

$$m \sum_{i=1}^{N} x_i^2 + c \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x_i y_i \quad \text{— (1)}$$

$$\frac{\partial L}{\partial c} = \sum_{i=1}^{N} y_i - m \sum_{i=1}^{N} x_i - Nc = 0$$

$$m \sum_{i=1}^{N} x_i + Nc = \sum_{i=1}^{N} y_i \quad \text{— (II)}$$

Samajh lo.

Representing the two equations in Matrix format

$$m\sum_{i=1}^{N} x_i^2 + c\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x_i y_i$$

$$m\sum_{i=1}^{N} x_i + c\sum_{i=1}^{N} 1 = \sum_{i=1}^{N} y_i$$

$$\begin{bmatrix} \sum_{i=1}^{N} x_i^2 & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} x_i y_i \\ \sum_{i=1}^{N} y_i \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix}}$$

$ax + by = e \leftarrow$

$cx + dy = f \leftarrow$

2 equation

Representing the two equations in Matrix format

(done)

Linear Regression

Gate (2)

A set of observations of independent variable (x) and the corresponding dependent variable (y) is given below.

x	5	2	4	3
y	16	10	13	12

So $a = 6.1$
Find b

$y = mx + c$

$m = \frac{Cov(x,y)}{Var(x)}$, $c = \bar{y} - m\bar{x}$

Based on the data, the coefficient a of the linear regression model

$y = a + bx$ is estimated as 6.1

The coefficient b is ____________ . (round off to one decimal place)

So $\bar{y} = \dfrac{16+10+13+12}{4} = 12.75$

$\bar{x} = \dfrac{5+2+4+3}{4} = 3.5$

So $C = \bar{y} - m\bar{x}$

$6.1 = 12.75 - m \times 3.5$

$m = 1.9$

Correlation Coefficient $\Rightarrow$

Correlation coef of two RV

$$\left(\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \, \sigma_Y} \right) = \frac{Covariance(X,Y)}{\text{Std dev of } X \cdot \text{Std dev of } Y}$$

Linear Regression

$Var = (Std\ dev)^2$

For a bivariate data set on (x, y), if the means, standard deviations and correlation coefficient are

$\bar{x} = 1.0$, $\bar{y} = 2.0$, $s_x = 3.0$, $s_y = 9.0$, $r = 0.8$

Then the regression line of y on x is:

1. $y = 1 + 2.4(x - 1)$
   $= 2.4x - 1.4$
2. $y = 2 + 0.\cancel{2}7(x - 1)$
3. $y = 2 + 2.4(x - 1)$ ✓
   $2.4x - .4$
4. $y = 1 + 0.27(x - 2)$

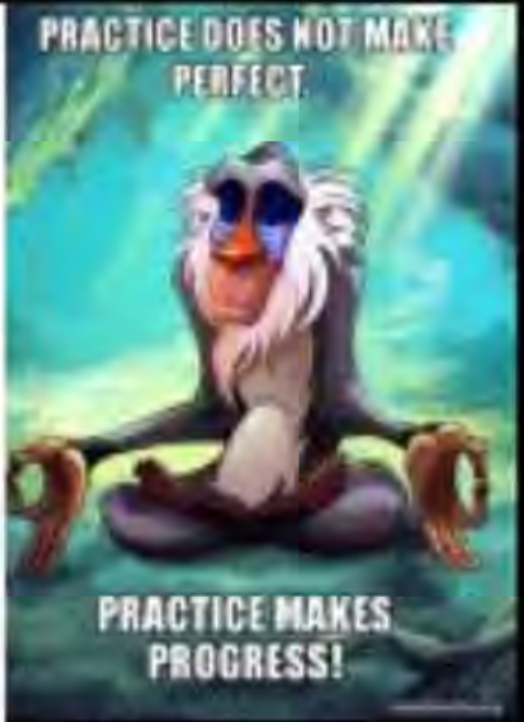
$\bar{x} = 1$

$\bar{y} = 2$

$\sigma_x = 3.0$ ✓

$\sigma_y = 9$ ✓

$\rho_{xy} = .8$ ✓



Regression line

$y = mx + c$

$m = \dfrac{Cov(x,y)}{Var(x)}$

$\rho_{xy} = \dfrac{Cov(x,y)}{\sigma_x \sigma_y}$

$.8 = \dfrac{Cov(x,y)}{3 \times 9}$

So $Cov(x,y) = 3 \times 9 \times \cdot 8$

$$m = \frac{Cov(x,y)}{Var(x)} = \frac{3 \times 9 \times \cdot 8}{(3)^2} = 2 \cdot 4$$

$$C = \bar{y} - m\bar{x}$$

$$C = 2 - 2 \cdot 4 \times 1$$

$$C = - \cdot 4$$

So $y = mx + c$

$\Rightarrow \left(y = 2 \cdot 4x - \cdot 4 \right)$

Linear Regression

In the regression model (y = a + bx) where $\bar{x}$ = 2.50, $\bar{y}$ = 5.50 and a = 1.50 ($\bar{x}$ and $\bar{y}$ denote mean of variables x and y and a is a constant), which one of the following values of parameter 'b' of the model is correct?

$$y = a + bx$$

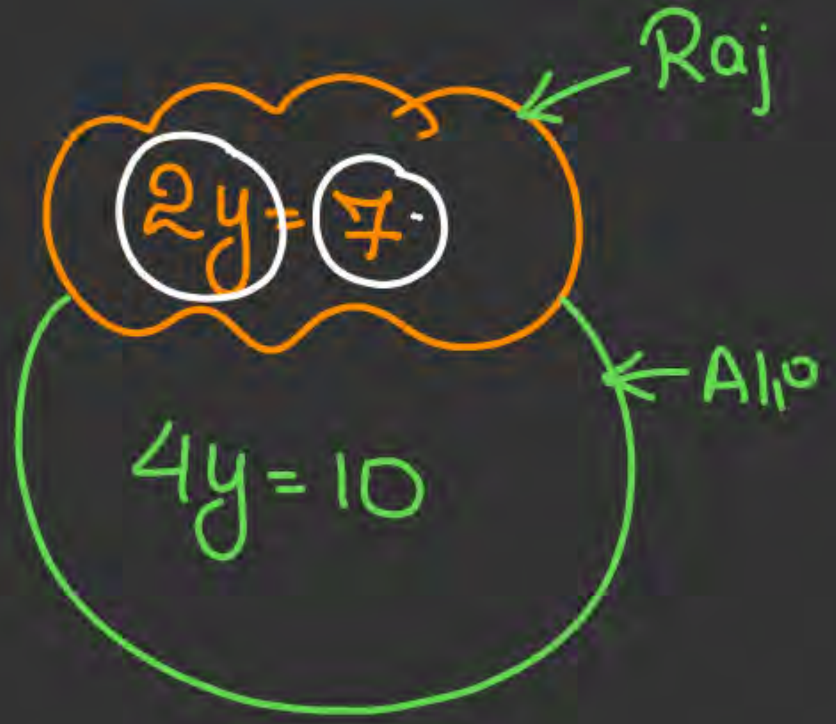
$$a \to c, \quad b \to m$$

$$c = \bar{y} - m\bar{x}$$

$$1.5 = 5.5 - m(2.5)$$

$$m = 1.6$$

1. 1.75
2. 1.60
3. 2.00
4. 2.50



Let we take value of $y = \hat{y}$
Such that error in both case of
Ali/Raj is minimized

$$\text{Loss} = \sum (\text{actual} - \text{Pred})^2$$
$$= (7 - 2\hat{y})^2 + (10 - 4\hat{y})^2$$

$\dfrac{\partial L}{\partial \hat{y}} = 0$ we get $\hat{y}$

There is no value of x that can simultaneously satisfy both the given equations. Therefore, find the 'least squares error' solution to the two equations, i.e., find the value of x that minimizes the sum of squares of the errors in the two equations. ________

$2x = 3$

$4x = 1$

$2x = 3$
$4x = 1$

$\Rightarrow$ (error)

Let $\hat{x}$ is the value of x that give min error $(2\hat{x}-3)^2 + (4\hat{x}-1)^2 \Rightarrow L$

3 = actual value

$2\hat{x}$ = Predicted "

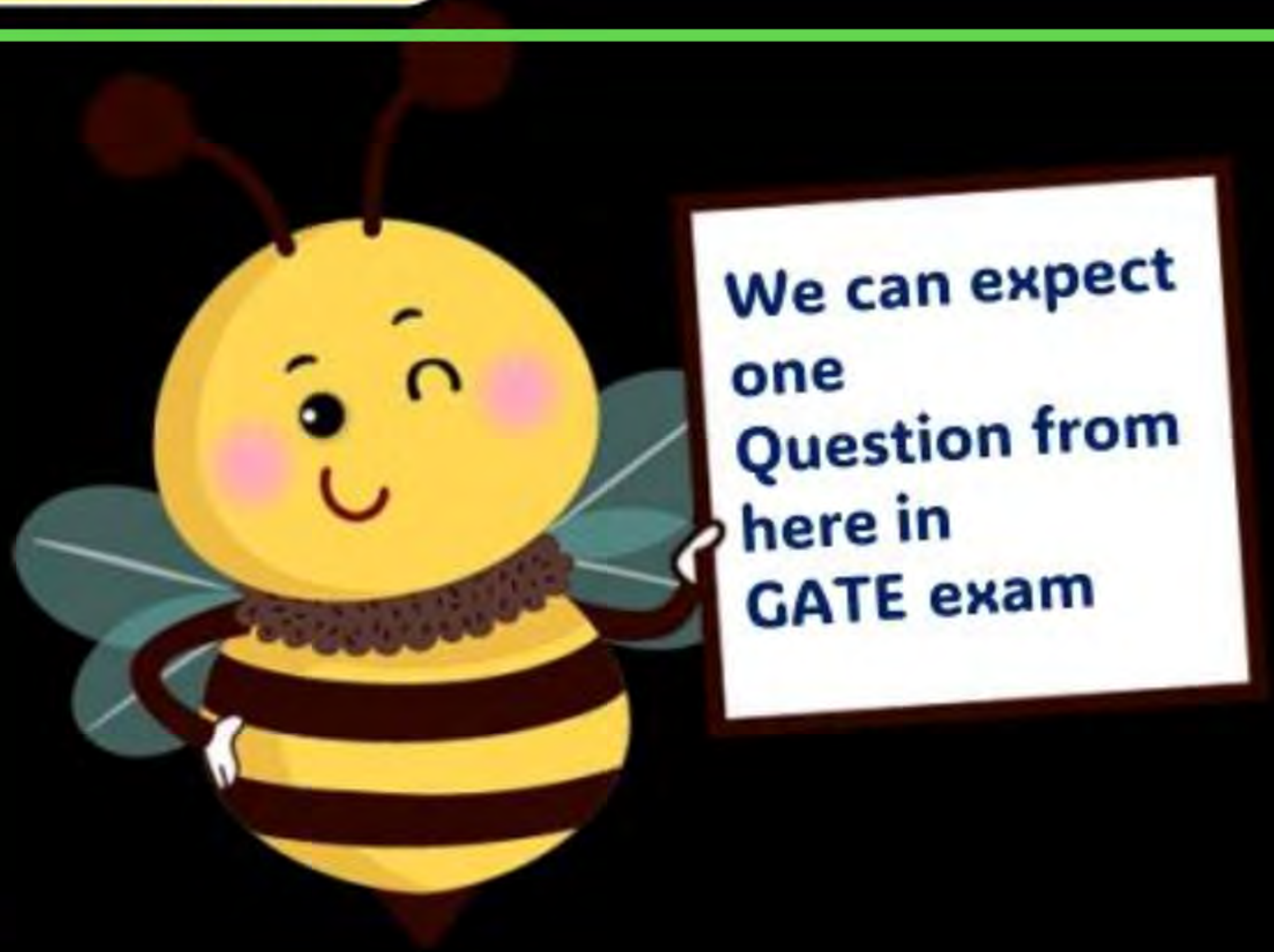
1 = actual value

$4\hat{x}$ = Predicted "

$$\frac{\partial L}{\partial \hat{x}} = 2(2\hat{x}-3)\times 2 + 2(4x-1)\times 4 = 0$$

$$8\hat{x} - 12 + 32\hat{x} - 8 = 0$$

$$40\hat{x} = 20, \quad \hat{x} = 1/2 \checkmark$$


We can expect
one
Question from
here in
GATE exam

Linear Regression

Considering data of 2 Dimensions

Attributes, Features, Dimensions...

2 input in each data

This data →

2 dimension in data

Income (LPA)	Age	Sale of I-Phone (in a month)
20	30	300
50	40	400
70	50	300
We have N Data points		

→ label "y"

Now the input data is 2 D (age and income)

How to write the 2 D inputs ??

Single dimension data

$(x_1, y_1) \leftarrow$ dataPoint 1

$(x_2, y_2) \leftarrow$ dataPoint 2

⋮

→ Suffix values Show datapoint

$y = mx + c$

2D data

Superscript ⇒ dimension
Subscript ⇒ datapoint No.

$(x_1^1, x_1^2, y_1) \leftarrow$ data No 1

$(x_2^1, x_2^2, y_2) \leftarrow$ data No 2

$(x_3^1, x_3^2, y_3) \leftarrow$ data No 3.

⋮

Here we have 2D ⇒ y will depend on $1^{st}/2^{nd}$ dimension ⇒ L.R ⇒ $y = \beta_0 + \beta_1 x^1 + \beta_2 x^2$

Linear Regression

Linear model will have 3 number of parameters

So in 2D $\Rightarrow$

$\Rightarrow \left(y = \underbrace{\beta_0 + \beta_1 x^1 + \beta_2 x^2}\right)$ 3 parameters i.e $\beta_0, \beta_1, \beta_2$.

datapoint $\underbrace{(x_i^1, x_i^2, y_i)}$ → Predicted Value $\hat{y}_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2$

actual Value

The optimisation method and equation will be ...

So RSS ⇒ Loss fxn

$$\boxed{\mathcal{L} = \sum_{i=1}^{N}\left(y_i - \left(\beta_0 + \beta_1 x_i^{1d} + \beta_2 x_i^{2d}\right)\right)^2}$$

let us have 3 points in data

$$\frac{\partial L}{\partial \beta_0} = \sum_{i=1}^{3} y_i - \beta_0 \sum_{i=1}^{3} 1 - \beta_1 \sum_{i=1}^{3} x_i^{1d} - \beta_2 \sum_{i=1}^{3} x_i^{2d} = 0$$

$$\frac{\partial L}{\partial \beta_1} = \sum_{i=1}^{3} x_i^{1d} y_i - \beta_0 \sum_{i=1}^{3} x_i^{1d} - \beta_1 \sum_{i=1}^{3} (x_i^{1d})^2 - \beta_2 \sum_{i=1}^{3} x_i^{2d} \cdot x_i^{1d} = 0$$

$$\frac{\partial L}{\partial \beta_2} = \sum_{i=1}^{3} x_i^{2d} y_i - \beta_0 \sum_{i=1}^{3} x_i^{2d} - \beta_1 \sum x_i^{1d} \cdot x_i^{2} - \beta_2 \sum (x_i^{2d})^2 = 0$$

Linear Regression

The optimisation method and equation will be ...

So
$$\beta_0 \sum_{i=1}^{3} 1 + \beta_1 \sum_{i=1}^{3} x_i^{1\,d} + \beta_2 \sum_{i=1}^{3} x_i^{2\,d} = \sum_{i=1}^{3} y_i^{\circ}$$

$$\beta_0 \sum_{i=1}^{3} x_i^{1} + \beta_1 \sum (x_i^{1})^2 + \beta_2 \sum x_i^2 x_i^1 = \sum_{i=1}^{3} x_i^1 y_i^{\circ}$$

$$\beta_0 \sum_{i=1}^{3} x_i^{2} + \beta_1 \sum x_i^1 x_i^2 + \beta_2 \sum (x_i^{2})^2 = \sum_{i=1}^{3} x_i^2 \cdot y_i^{\circ}$$

done

$$\begin{bmatrix} \sum 1 & \sum x_i^1 & \sum x_i^2 \\ \sum x_i^1 & \sum (x_i^1)^2 & \sum x_i^2 x_i^1 \\ \sum x_i^2 & \sum x_i^1 x_i^2 & \sum (x_i^2)^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \sum y_i^{\circ} \\ \sum x_i^1 y_i^{\circ} \\ \sum x_i^2 y_i^{\circ} \end{bmatrix}$$

The optimisation method and equation will be ...

data Representation

We know that $y = \beta_0 + \beta_1 x^1 + \beta_2 x^2$

data point $\Rightarrow \begin{bmatrix} x_i^1 & x_i^2 \end{bmatrix}$

Any data point 2D

$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$

To find $y \Rightarrow \begin{bmatrix} x_i^1 & x_i^2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$

$\Rightarrow \beta_1 x_i^1 + \beta_2 x_i^2$

But add β_0

$y = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2$

Linear Regression

So we add extra '1' in data

The representation of 2-dimensional data in matrix format

So datapoint $\left(\underset{\sim}{1} \quad x_i^1 \quad x_i^2\right) \longleftrightarrow \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$

$$\left(1 \quad x_i^1 \quad x_i^2\right) \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \Rightarrow \left(\beta_0 + x_i^1 \beta_1 + x_i^2 \beta_2\right)$$

So in matrix form $\Rightarrow$ datapoint $\begin{bmatrix} 1 & x_i^1 & x_i^2 \end{bmatrix}$ $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$

Linear Regression

The representation of 2-dimensional data in matrix format

So we had 3 points $\Rightarrow X \Rightarrow \begin{bmatrix} 1 & x_1^1 & x_1^2 \\ 1 & x_2^1 & x_2^2 \\ 1 & x_3^1 & x_3^2 \end{bmatrix}$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$\rightarrow$ actual Y values of data

$$X^T = \begin{bmatrix} 1 & 1 & 1 \\ x_1^1 & x_2^1 & x_3^1 \\ x_1^2 & x_2^2 & x_3^2 \end{bmatrix}$$

$$X^TY \Rightarrow \begin{bmatrix} \sum_{i=1}^{3} y_i \\ \sum_{i=1}^{3} x_i^1 y_i^0 \\ \sum_{i=1}^{3} x_i^2 y_i \end{bmatrix}$$

Linear Regression

The representation of 2-dimensional data in matrix format

So we had 3 points $\Rightarrow X \Rightarrow \begin{bmatrix} 1 & x_1^1 & x_1^2 \\ 1 & x_2^1 & x_2^2 \\ 1 & x_3^1 & x_3^2 \end{bmatrix}$

$$X^T = \begin{bmatrix} 1 & 1 & 1 \\ x_1^1 & x_2^1 & x_3^1 \\ x_1^2 & x_2^2 & x_3^2 \end{bmatrix}$$

$$X^T X \Rightarrow \begin{bmatrix} 1 & 1 & 1 \\ x_1^1 & x_2^1 & x_3^1 \\ x_1^2 & x_2^2 & x_3^2 \end{bmatrix} \begin{bmatrix} 1 & x_1^1 & x_1^2 \\ 1 & x_2^1 & x_2^2 \\ 1 & x_3^1 & x_3^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{3} 1 & \sum_{i=1}^{3} x_i^1 & \sum_{i=1}^{3} x_i^2 \\ \sum_{i=1}^{3} x_i^1 & \sum_{i=1}^{3} (x_i^1)^2 & \sum_{i=1}^{3} x_i^1 x_i^2 \\ \sum_{i=1}^{3} x_i^2 & \sum_{i=1}^{3} (x_i^2)(x_i^1) & \sum_{i=1}^{3} (x_i^2)^2 \end{bmatrix}$$

Linear Regression

Bas itna hi Yaad Karo

The representation of 2-dimensional data in matrix format

Summarise

- $X \Rightarrow$ Show data
  2D data

$$\begin{bmatrix} 1 & x_1^1 & x_1^2 \\ 1 & x_2^1 & x_2^2 \\ 1 & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{bmatrix}_{N\times 3}$$

N data points

actual y values

- $Y \Rightarrow \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ \vdots \end{bmatrix}_{N\times 1}$

- $\beta \Rightarrow \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}_{3\times 1}$

Linear Regression

Bas itna hi Yaad.

The representation of 2-dimensional data in matrix format

$$\boxed{(X^TX)\beta = (X^TY)}$$ ⇐ Resut of the min of loss function

So we can find β's [i.e coefficient of linear regression]

$$\beta = (X^TX)^{-1}(X^TY)$$ ⇐

Linear Regression

The final expression of Beta in 2 D case ...

data of D number of dimension

$$X = \begin{bmatrix} 1 & x_1^1 & x_1^2 & \cdots & x_1^D \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}_{N \times (D+1)}$$ N points

Formula ⇒

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

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

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_D \end{bmatrix}_{(D+1) \times 1}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$ actual y

The final expression of Beta in 2 D case ...

2D data			
Temp	Pressure	Rain	Wind speed

$$\text{Rain} = \beta_0 + \beta_1 P + \beta_2 T$$

$$\text{Wind speed} = \beta_0 + \beta_1 P + \beta_2 T$$

The final expression of derivative of L by Beta ...

Linear Regression

Now lets extend the whole into D dimensional data

The loss function for D dimensions case

Loss function in Matrix Form

Considering data of P Dimensions

The loss function for P dimensions case

Loss function in Matrix Form

We do partial differentiation in terms of all variables to get the optimized variable values

Considering data of P Dimensions

R-squared in Regression Analysis in Machine Learning

$$R^2 = 1 - \frac{RSS}{TSS}$$

R^2 = coefficient of determination

RSS = sum of squares of residuals

TSS = total sum of squares

Linear Regression

Considering data of P Dimensions

R-squared in Regression Analysis in Machine Learning

$$RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

RSS = residual sum of squares

y_i = i^th value of the variable to be predicted

$f(x_i)$ = predicted value of y_i

n = upper limit of summation

Considering data of P Dimensions

R-squared in Regression Analysis in Machine Learning

$$\mathrm{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

TSS = total sum of squares

n = number of observations

y_i = value in a sample

$\bar{y}$ = mean value of a sample

Considering data of P Dimensions

R-squared in Regression Analysis in Machine Learning

- The most important thing we do after making any model is evaluating the model.
- R-squared is a statistical measure that represents the goodness of fit of a regression model.
- The value of R-square lies between 0 to 1.
- Where we get R-square equals 1 when the model perfectly fits the data and there is no difference between the predicted value and actual value.
- However, we get R-square equals 0 when the model does not predict any variability in the model.

Linear Regression

Considering data of P Dimensions

R-squared in Regression Analysis in Machine Learning

- R-Squared (R^2 or the coefficient of determination) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable.
- The most common interpretation of r-squared is how well the regression model explains observed data. For example, an r-squared of 60% reveals that 60% of the variability observed in the target variable is explained by the regression model.

Considering data of P Dimensions

R-squared in Regression Analysis in Machine Learning

- The goodness of fit of regression models can be analyzed on the basis of the R-square method. The more the value of the r-square near 1, the better the model is.
- Note: The value of R-square can also be negative when the model fitted is worse than the average fitted model. .

Linear Regression

Considering data of P Dimensions

Adjusted R - Squares

- Adjusted R-Squared is an updated version of R-squared which takes account of the number of independent variables while calculating R-squared.
- n is the total number of observations in the data
- k is the number of independent variables (predictors) in the regression model

$$AdjustedR^2 = 1 - \frac{(1-R^2)\cdot(n-1)}{n-k-1}$$

2 mins Summary

Topic Lineare Reg.

Topic

Topic

Topic

Topic

THANK - YOU