Data Science and Artificial Intelligence

Machine Learning

Regression

Lecture No. 08

By- SIDDHARTH SABHARWAL SIR

Recap of Previous Lecture

Topic Multicollinearity

Topic VIF, Covariance matrix

Topic Assumption in LR

Topic

Topic

Topics to be Covered

Topic — Advantage & disadvantage of LR

Topic — Time & space complexity of LR.

Topic

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


FAILURE IS
A BRUISE, NOT
A TATTOO.
JON SINCLAIR

REVISION

$$VIF = \frac{1}{1-R^2}$$

VIF = Large : Multicollinearity exist

VIF = Small : No multicollinearity

Correlation matrix
- → diagonal elements = 1
- → Rest other are close to ±1 → Multi...
- other elements ≈ 0 → No Multi...

REVISION

Assumptions

→ Homoscedasticity ⇒ • Noise in data shd be independent of x.

→ No Multicollinearity

→ datapoints Independent

→ y and x shd have linear Relation

(Siddharth Sin PW)

Considering data of P Dimensions

Lets Practice

Based on the data provided below, answer questions from (7-10). We consider a function we wish to minimize. $J(w) = \frac{1}{10}\Sigma_i^5 (y^{(i)} - w_1 x^{(i)} - w_0)^2$ where the constants $x^{(i)}, y^{(i)}$ are provided in the table below

i	$x^{(i)}$	$y^{(i)}$
1	0	1
2	1	3
3	2	5
4	3	8
5	4	9

Dataset

7) The dimension of w is ________.

GAATE 2026 FODENGE

Considering data of P Dimensions

Lets Practice

8) Start with the initial guess of $[w_0, w_1] = [0, 0]$. Take the value of learning rate = 1. The value of w_0 after 4 iterations of gradient descent will be __________.

Space and Time Complexity of Linear Regression

Time Complexity ⇒

Training time Complexity → In LR during training we want to find model, Relationship, Coefficients

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

$$[A]_{m \times n} [B]_{n \times p} = [R]_{m \times p}$$

$$
\begin{array}{c|c}
 & 1 \; 2 \; 3 \; \cdots \; n \\
\begin{array}{c}1\\2\\3\\\vdots\\m\end{array} & \begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix}_{m\times n}
\end{array}
\begin{array}{c|c}
 & 1 \; 2 \; 3 \; \cdots \; p \\
\begin{array}{c}1\\2\\3\\\vdots\\n\end{array} & \begin{bmatrix} & & \\ & B & \\ & & \end{bmatrix}_{n\times p}
\end{array}
=
\begin{array}{c|c}
 & 1 \; 2 \; \cdots \; p \\
\begin{array}{c}1\\2\\\vdots\\m\end{array} & \begin{bmatrix} \bigcirc & & \\ & & \\ & & \end{bmatrix}_{m\times p}
\end{array}
$$

- To generate each term we multiply n values
- Total $m \times p$ values

$\Rightarrow$ Total No of mult $\Rightarrow m \times n \times p$

Space and Time Complexity of Linear Regression

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}_{3\times3}$$

$$\left.\begin{aligned} \rightarrow M_{11} &= ei - fh \\ M_{12} &= di - fg \end{aligned}\right\}$$

• for each minor we need 2 mult

$$A^{-1} = \frac{1}{|A|}[\text{Adj}]$$

$\frac{1}{|A|}$ → (det of matrix A)

$$\text{Adj} = [\text{Cof}]^T$$

$$\rightarrow C_{ij} = (-1)^{i+j} M_{ij}$$

M_{ij} → for each cofactor we need 3 mult

M_{ij} → Minor

• Cofactor matrix will have 3×3 in size ⇒ 9 cofactors

$\left.\begin{aligned} &\Rightarrow 3\times3\times3 \\ &\Rightarrow 3^3 \end{aligned}\right\}$ No of mult for Cof. matrix

• 3×3 matrix inverse
⇒ we need 3×3×3
No of mult ⇒ 3^3

Space and Time Complexity of Linear Regression

So To find inverse of $\underbrace{k \times k}_{\text{Square}}$ matrix $\longrightarrow$ k^3 No of multiplication

N: data points

D: No of dimension

$X = [\ ]_{N \times P}$

$Y = [\ ]_{N \times 1}$

$\beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_D \end{bmatrix}_{P \times 1}$

$D + 1 = P$

$(X^T X) \Rightarrow (X^T)_{P \times N} (X)_{N \times P} = (X^T X)_{P \times P}$

$\hookrightarrow$ No of Mult $\Rightarrow P \times N \times P = NP^2$

$(X^T X)_{P \times P} \longrightarrow (X^T X)^{-1}$

$\hookrightarrow$ No of Mult $\Rightarrow P^3$

$X^T Y \Rightarrow (X^T)_{P \times N} (Y)_{N \times 1} \Rightarrow (X^T Y)_{P \times 1}$

$\hookrightarrow$ No of mult $\Rightarrow N \times P \times 1$

Train time Complexity.

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

$\hookrightarrow$ No of mult $- p^2$

Total No of mult $\Rightarrow (Np^2 + p^3 + Np + p^2)$

• So Time Complexity for Training $\Rightarrow$ Order of $O[p^3 + Np^2 + Np + p^2]$

* These are exact number of operations $\longrightarrow$ No
* So order off means $\Rightarrow$ Range of No of Calculations.

Space Complexity

Linear Regression

Training Data (x y) → Training → model β's

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

After training we only need to store model
⇒ i.e store the β values, β matrix
⇒ So we need to store $(D+1)$ values

⇒ Space Complexity : memory used to store model
⇒ Order of $(D+1)$
$O(D+1)$

Test time Complexity

Linear Regression

new data Point → Testing → $\hat{y}$ (Predicted value)

model eq $y = \beta_0 + \beta_1 x^1 + \beta_2 x^2 - - \beta_D x^D$

Now to find $\hat{y}$ for test Point, Simply Put values of x's in eq of model

$$\hat{y} = \beta_0 + \beta_1 x_t^1 + \beta_2 x_t^2 - - - \beta_D x_t^D$$ ← $\hat{y}$ Value

For Testing we need D number of mult → Test time Complexity = O(D)

Train time Comp: $O[p^3 + Np^2 + Np + p^2]$

Test time Complexity $O(D)$

Space Complexity $O(D+1)$

Disadvantage

1. underfit on the Non linear data

2. affected by outlier

3. If dimensions are dependent LR model Perform poorly.

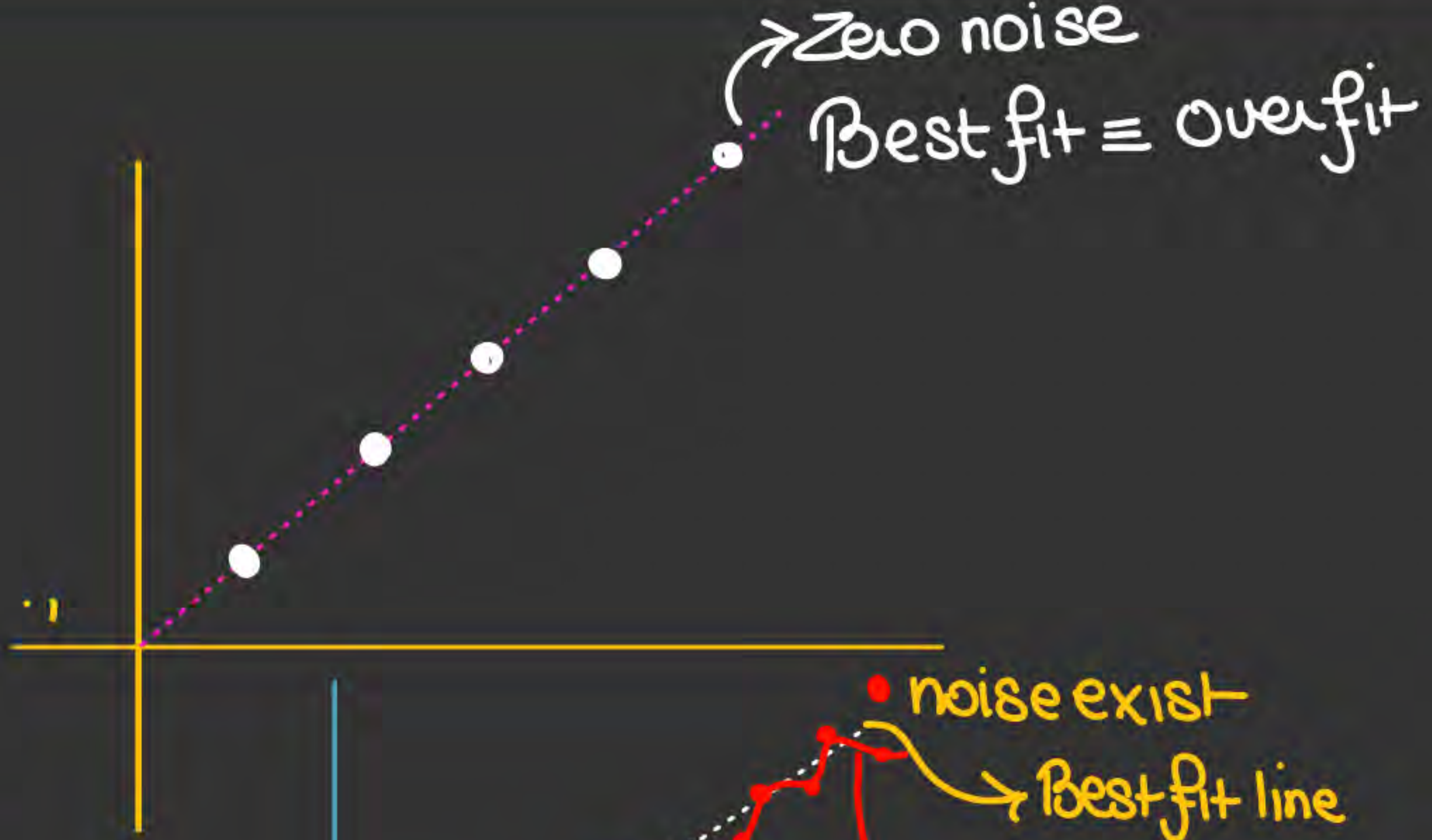
4. Data shd have Homo-Scedasticity, then only LR algo works.

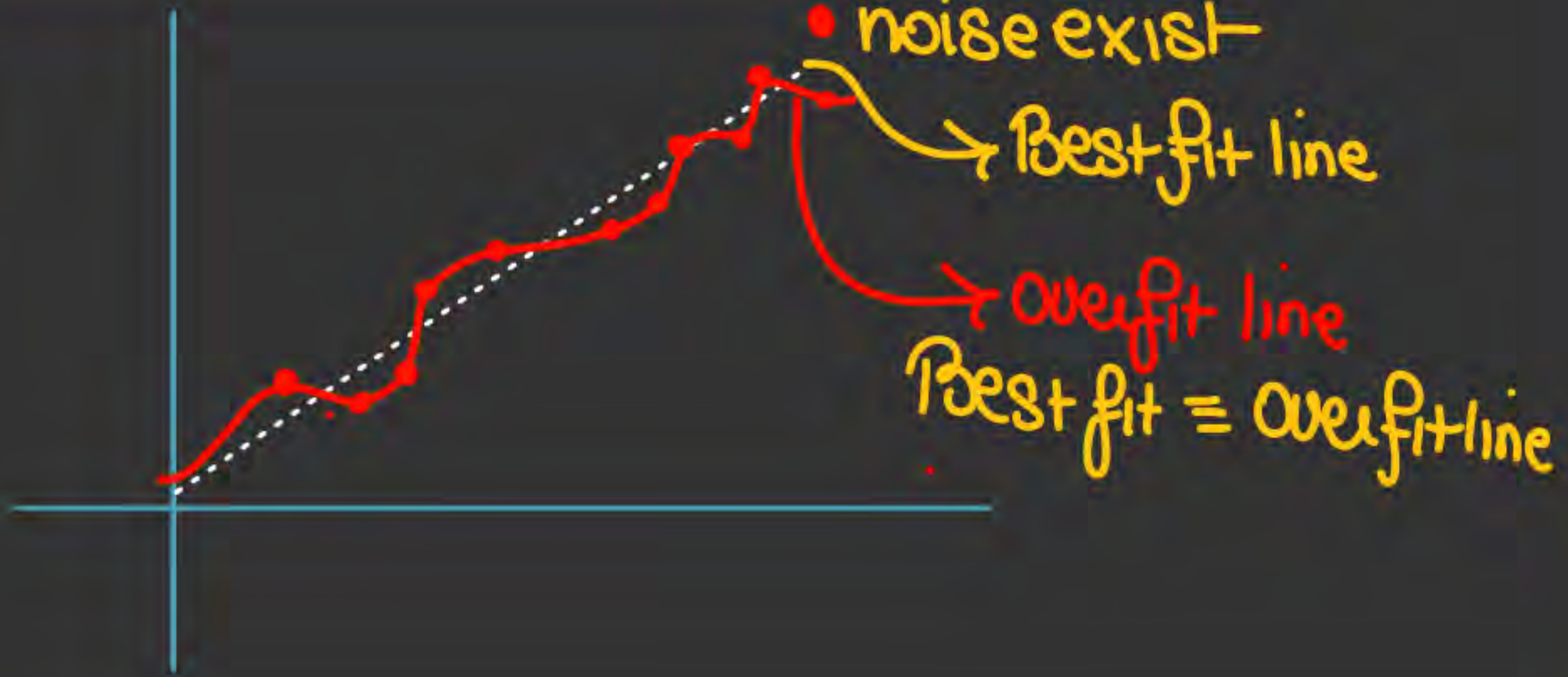
5. The LR algorithm Tend to overfit

→ data Ko Yaad

Rattu Tota

→ Badhiya data diya
not noisy → data Pattern





General Points ⇒

- So model that try to provide Best fit ⇒ less/not affected by noisy/outlier points
- model that try to overfit data ⇒ are affected by noisy/outlier points.

Linear Regression Only one Task ⇒ $L = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$

To (min L)

→ So algo has only one Task ⇒ min gap b/w y and $\hat{y}$

⇒ want to equate y and $\hat{y}$ ⇒ Tendency to overfit.

Linear Regression advantages

1) Interpretability is highest

2) Simplest algo

3) We got to know Importance of each dimension from β's.

Linear Regression

Why Linear Regression is Important

- The interpretability of linear regression is a notable strength.
- The model's equation provides clear coefficients that elucidate the impact of each independent variable on the dependent variable, facilitating a deeper understanding of the underlying dynamics.
- Its simplicity is a virtue, as linear regression is transparent, easy to implement, and serves as a foundational concept for more complex algorithms.

y and $\hat{y}$

- Overfit ⇒ Residual Plot → 0

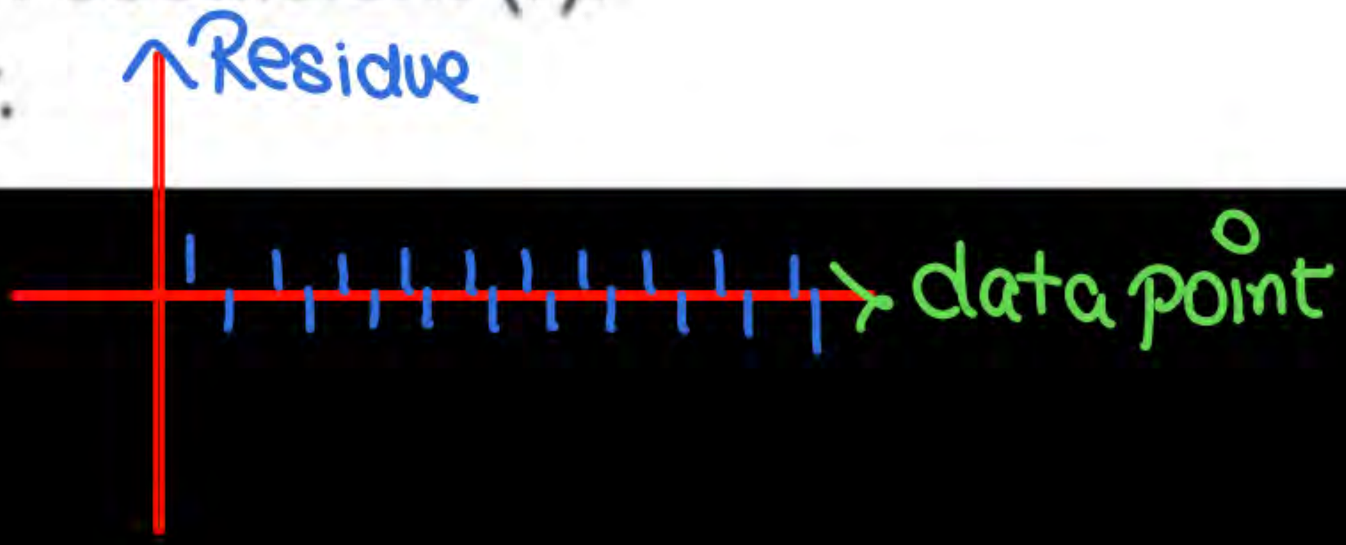
What is the purpose of the residual plot in multiple linear regression analysis?

A) To visualize the relationship between independent variables.

B) To check for homoscedasticity and the presence of outliers.

C) To calculate the correlation coefficient (r).

D) To assess multicollinearity.



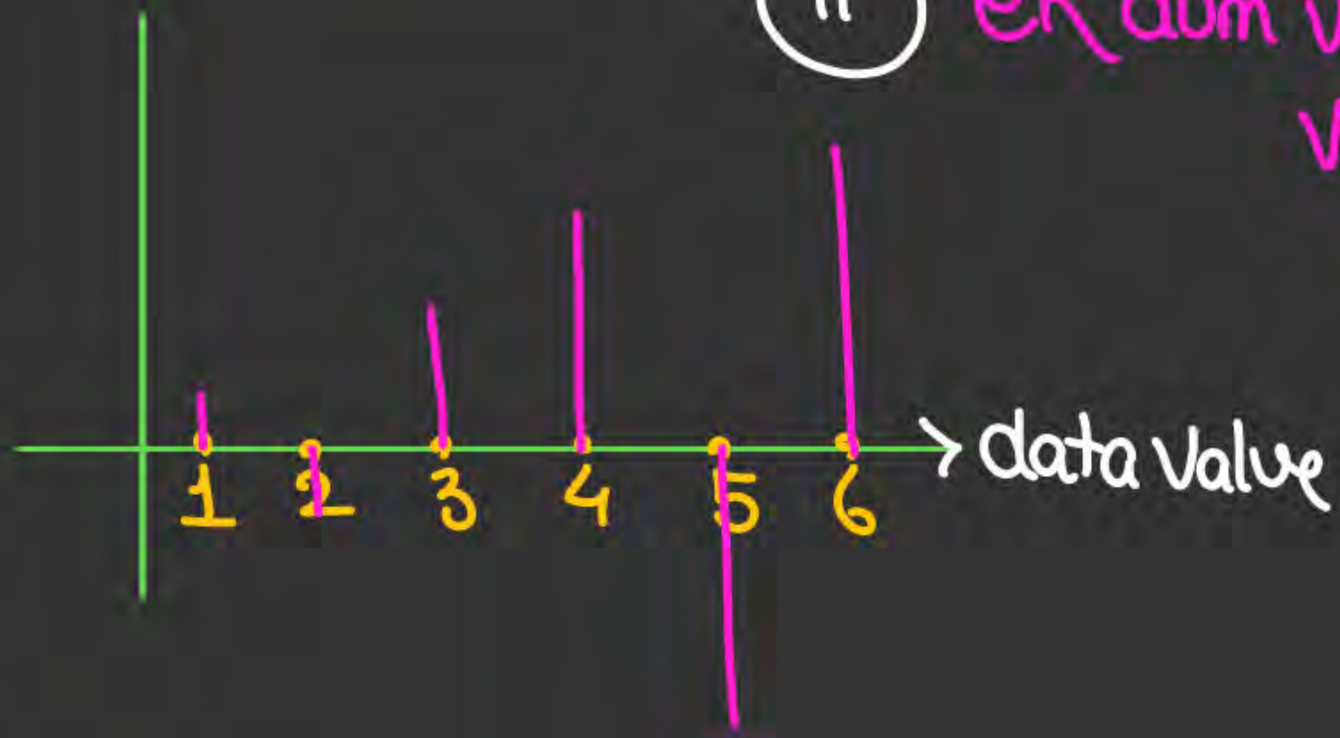
Residue ±ve

→ Similar Nature →

— Homoscedasiticity

Residual Plot ⇒ (i) ± ve equal small values = Homo...

(ii) ek dum values inc,
Value of Residue inc with x ⇒ Hetero...



What is the main purpose of the intercept term in a multiple linear regression model?

$y = \beta_0 + \beta_1 x^1 + --- \beta_D x^D$

~~A)~~ It represents the slope of the regression line.

all ind. variable $x^1, x^2, x^D = 0$

~~B)~~ It is used to control for multicollinearity.

$y = \beta_0$

C) It represents the expected value of the dependent variable when all independent variables are zero.

~~D)~~ It is not used in multiple linear regression.

Question 15: What is the purpose of the coefficient of determination (R-squared) in simple linear regression?

A. To determine the slope of the regression line
B. To measure the strength of the linear relationship
C. To calculate the p-value of the regression
D. To identify outliers in the dataset

- In LR, R^2 find goodness of model

Question 2: What does the coefficient of determination (R-squared) measure in multiple linear regression?

x

A. The correlation between predictor variables
B. The percentage of variance in the dependent variable explained by the model
C. The significance of the intercept term
D. The number of predictor variables in the model

Question 12: In simple linear regression, which variable is considered the independent variable?

x

A. The variable being predicted → y

B. The response variable → y

C. The predictor variable

D. There is no independent variable in simple linear regression negatively Correlated $\Rightarrow$ if $x\uparrow, R\downarrow$ OR $x\downarrow, R\uparrow$

Positively Correlated if $x\uparrow, R\uparrow$ $x\downarrow, R\downarrow$

Question 20: Which of the following statements is true regarding the residual plot in simple linear regression?

A. Residuals should exhibit a clear linear pattern.
B. Residuals should be randomly scattered around the horizontal line.
C. Residuals should be negatively correlated with the predictor variable.
D. Residuals should have a positive correlation with the dependent variable.

Heteroscedasticity

X_1	X_2	X_3		X_n	Y
a	b	c	..	d	y_1
e	f	g	..	h	y_2
i	j	K	-.	L	y_3
:					:

-ve show if X_1 inc Y dec

5.FOr a give N independent input variables (X1,X2.... Xn) and dependent (target) variable Y a linear regression is fitted for the best fit line using least square error on this data. The correlation coefficient for one of it's variable(Say X1) with Y is -0.97. Which of the following is true for X1? ?

- A) ~~Relation between the X1 and Y is weak~~
- B) Relation between the X1 and Y is strong ✓
- C) Relation between the X1 and Y is neutral
- D) Correlation does not imply relationship

$\rho \Rightarrow -1$ to 1

$0 \approx$ uncorrelated

$\pm 1 \Rightarrow$ highly correlated

$V_1 V_2 \rightarrow$ neither uncorrelated
nor highly correlated

6.Given below characteristics which of the following option is the correct for Pearson correlation between V1 and V2? If you are given the two variables V1 and V2 and they are following below two characteristics. 1. If V1 increases then V2 also increases 2. If V1 decreases then V2 behavior is unknown ?

- A) Pearson correlation will be close to 1
- B) Pearson correlation will be close to -1
- C) Pearson correlation will be close to 0 $\rightarrow$ Uncorrelated.
- D) None of these

2) In regression analysis, the variable that is being predicted is;

a) the independent variable

b) $\boxed{\textbf{the dependent variable}}$

c) usually denoted by x

d) usually denoted by r

6) Least square method calculates the best-fitting line for the observed data by minimizing the sum of the squares of the _______ deviations.

a) **Vertical**
b) Horizontal
c) Both of these
d) None of these

7) Which one is the least square method formula;

a) $\min \sum (y_i - \hat{y}_i)^2$

H.W b) $\min \sum (\hat{y}_i - y_i)$

c) $\min \sum (y_i - \hat{y}_i)^2$

d) $\min \sum (y_i - \hat{y}_i)$

13) Below you are given a summary of the output from a simple linear regression analysis from a sample of size 15, SSR=100, SST = 152. The coefficient of determination is;

$$R^2 = 1 - \frac{RSS}{TSS} \Rightarrow .3421$$

10) **A residual is defined as**

a) The difference between the actual Y values and the mean of Y.

b) The difference between the actual Y values and the predicted Y values.

c) The predicted value of Y for the average X value.

d) The square root of the slope.

H.W

11) If the regression equation is equal to y=23.6−54.2x, then 23.6 is the _____ while -54.2 is the _____ of the regression line.

a) Slope, intercept

b) Slope, regression coefficient

c) Intercept, slope

d) Radius, intercept

Q8. Suppose we have N independent variables (X1, X2... Xn) and Y's dependent variable.

Now Imagine that you are applying linear regression by fitting the best-fit line using the least square error on this data. You found that the correlation coefficient for one of its variables (Say X1) with Y is -0.95.

H.W

Which of the following is true for X1?

A) Relation between the X1 and Y is weak

B) Relation between the X1 and Y is strong

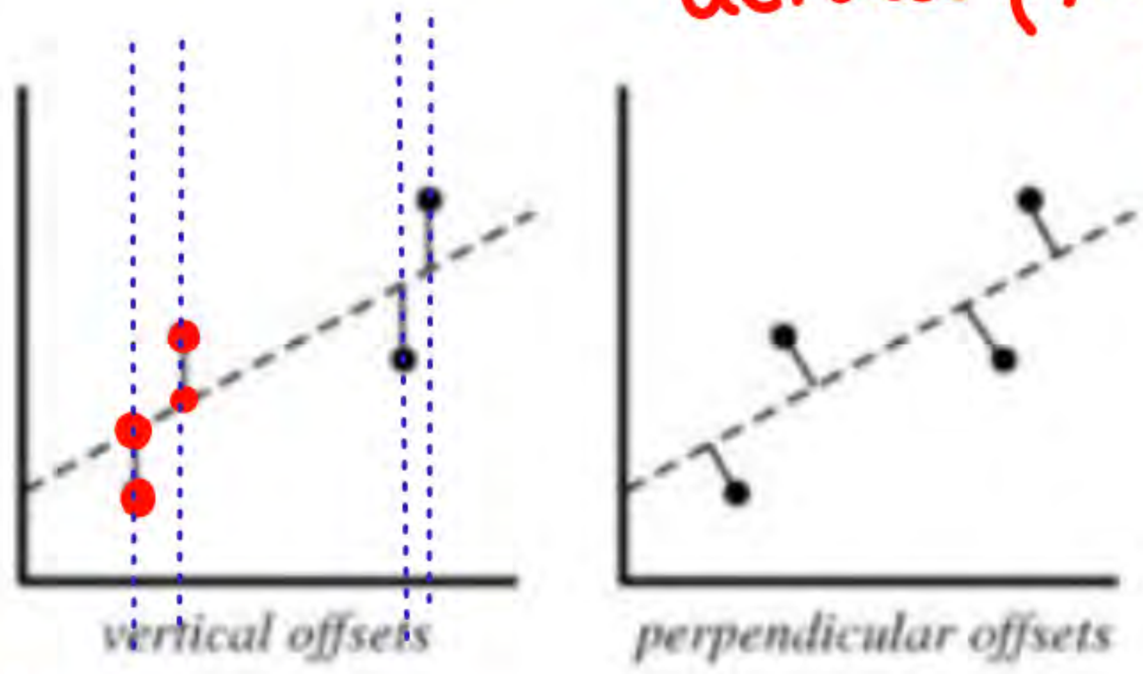
C) Relation between the X1 and Y is neutral

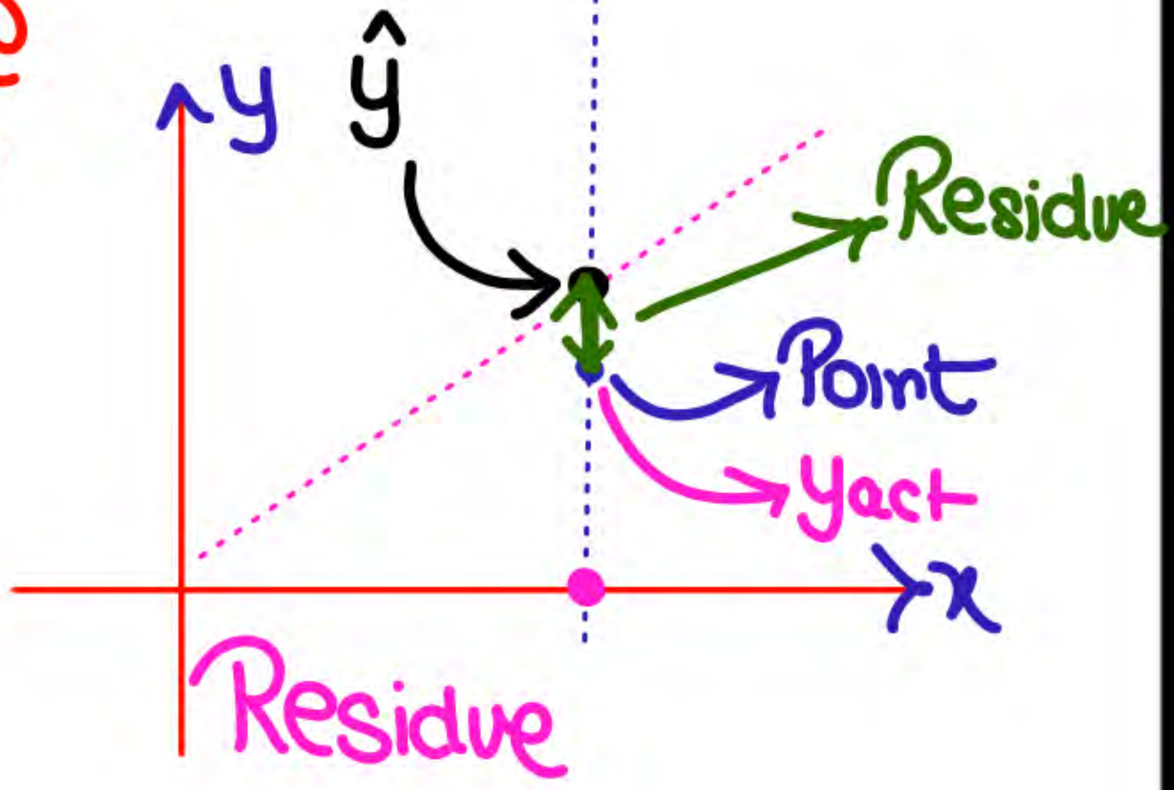
D) Correlation can't judge the relationship

Solution: (B)

Q11. Suppose the horizontal axis is an independent variable and the vertical axis is a dependent variable. Which of the following offsets do we use in linear regression's least square line fit?

actual & Predicted @ Same x





Vertical Offset.

B) Perpendicular offset

C) Both, depending on the situation

D) None of above

Q12. True- False: Overfitting is more likely when you have a huge amount of data to train.

H.W

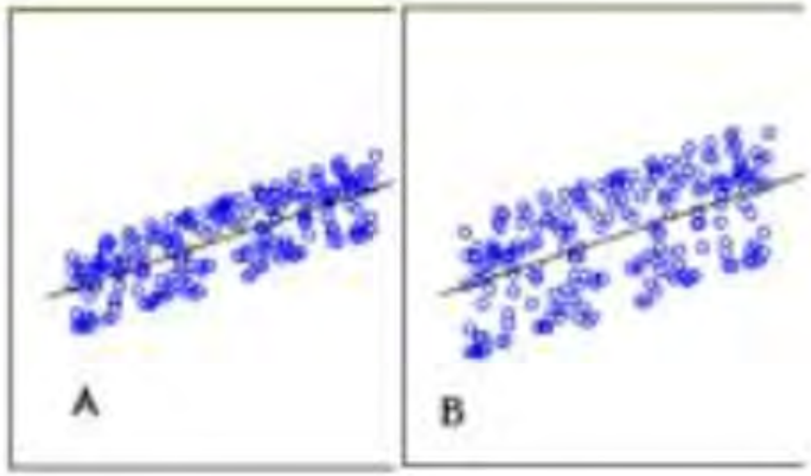
A) TRUE

B) FALSE

Solution: (B)

Q14. Which of the following statement is true about the sum of residuals of A and B?

Below graphs show two fitted regression lines (A & B) on randomly generated data. Now, I want to find the sum of residuals in both cases, A and B.



H.W

A) A has a higher sum of residuals than B

B) A has a lower sum of residual than B

C) Both have the same sum of residuals

D) None of these

Q18. Which of the following statement is true about outliers in Linear regression?

H.W

A) Linear regression is sensitive to outliers

B) Linear regression is not sensitive to outliers

C) Can't say

D) None of these

Q19. Suppose you plotted a scatter plot between the residuals and predicted values in linear regression and found a relationship between them. Which of the following conclusion do you make about this situation?

H.W

A) Since there is a relationship means our model is not good

B) Since there is a relationship means our model is good

C) Can't say

D) None of these

Suppose that you have a dataset D1 and you design a linear model of degree 3 polynomial and find that the training and testing error is "0" or, in other words, it perfectly fits the data.

Q20. What will happen when you fit a degree 4 polynomial in linear regression?

H.W

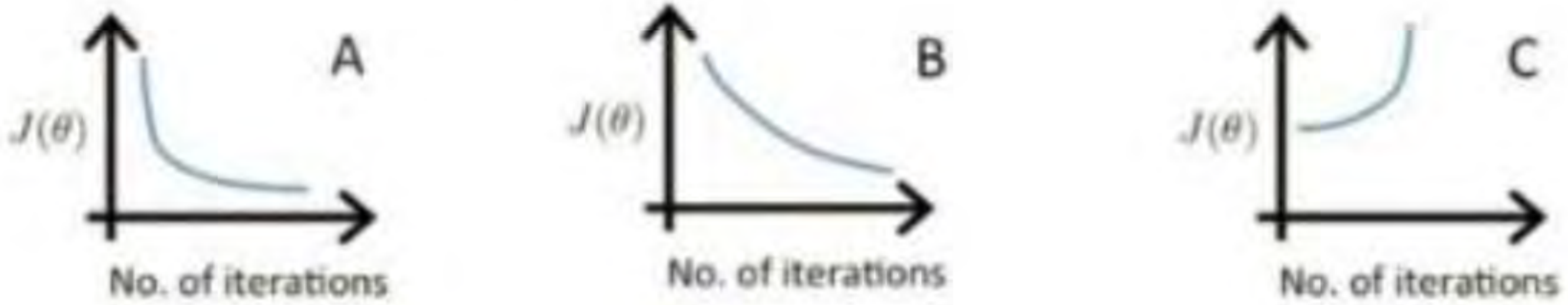
A) There is a high chance that degree 4 polynomial will overfit the data

B) There is a high chance that degree 4 polynomial will underfit the data

C) Can't say

D) None of these

Below are three graphs, A, B, and C, between the cost function and the number of iterations, l1, l2, and l3, respectively.



H.W

Q23. Suppose l1, l2, and l3 are the three learning rates for A, B, and C, respectively. Which of the following is true about l1,l2, and l3?

A) l2 < l1 < l3

B) l1 > l2 > l3

C) l1 = l2 = l3

D) None of these

Considering data of P Dimensions

Lets Practice

Based on the data provided below, answer questions from (7-10). We consider a function we wish to minimize. $J(w) = \frac{1}{10}\Sigma_i^5(y^{(i)} - w_1 x^{(i)} - w_0)^2$ where the constants $x^{(i)}$, $y^{(i)}$ are provided in the table below

i	$x^{(i)}$	$y^{(i)}$
1	0	1.4822
2	0.25	1.8165
3	0.50	1.9171
4	0.75	2.3930
5	1.00	2.5826

Dataset

7) The dimension of w is ___________.

THANK - YOU