- SIMPLE LINEAR REGRESSION

Q1. What is Simple Linear Regression?

Ans. Simple Linear Regression is a method used to find the relationship between one independent variable (X) and one dependent variable (Y) using a straight line.

Example: Predicting marks (Y) based on study hours (X).

Q2. Key assumptions of Simple Linear Regression Relationship between X and Y is linear

Ans. Errors are independent

Errors have constant variance (no heteroscedasticity)

Errors are normally distributed

No extreme outliers

Q3. What does coefficient m represent in Y = mX + c?

Ans. m = slope

It shows how much Y changes when X increases by 1 unit

If m = 2 → when X increases by 1, Y increases by 2

Q4. What does intercept c represent?

Ans. c = value of Y when X = 0

It is the starting point of the line

Q5. How do we calculate slope ? m

∑ (X − X ˉ) (Y − Y ˉ) ∑ (X − X ˉ) 2 m= ∑(X− X ˉ) 2

∑(X− X ˉ)(Y− Y ˉ)

In simple words: Slope shows the direction and strength of the relationship.

Q6. Purpose of Least Squares Method

Ans. It finds the best fitting line

It minimizes the sum of squared errors (difference between actual and predicted values)

Q7. Interpretation of R^2 (Coefficient of Determination)

Ans. R^2 shows how much variation in Y is explained by X

Value lies between 0 and 1

$R^2 = 0.80 \rightarrow 80\%$ variation explained

◆ MULTIPLE LINEAR REGRESSION

Q8. What is Multiple Linear Regression?

Ans. It studies the relationship between one dependent variable and two or more independent variables.

Example: Salary = experience + education + skills

Q9. Difference between Simple & Multiple Linear Regression

Ans. Simple Multiple One X variable Two or more X variables $Y = mX + c$ $Y = b_1X_1 + b_2X_2 + c$

Q10. Key assumptions of Multiple Linear Regression

Ans.

Linear relationship

No multicollinearity

Homoscedasticity

Errors are independent

Normal distribution of errors

Q11. What is heteroscedasticity?

Ans. When variance of errors is not constant

It leads to unreliable results

Causes wrong standard errors and tests

Q12. How to improve model with high multicollinearity?

Ans. Remove highly correlated variables

Use Ridge or Lasso regression

Use PCA

Combine variables

Q13. Techniques to transform categorical variables

Ans. One-hot encoding

Label encoding

Dummy variables

Q14. Role of interaction terms

Ans.Shows how two variables together affect Y

Example: education × experience

Q15. Intercept interpretation: Simple vs Multiple

Ans. Simple: Y when X = 0

Multiple: Y when all X variables = 0

Q16. Significance of slope in regression

Ans. Shows direction (+/–) and strength

Affects prediction accuracy

Q17. How does intercept give context?

Ans. It gives a baseline value

Helps understand where prediction starts

Q18. Limitations of using R^2 alone

Ans. Cannot detect overfitting

Always increases with more variables

Doesn't show causation

Q19. Interpretation of large standard error

Ans. Coefficient estimate is not reliable

Variable may be insignificant

Q20. Identifying heteroscedasticity in residual plots

Ans. Pattern looks like cone or fan shape

Important because it affects model validity

Q21. High R^2 but low adjusted R^2 means

Ans. Too many irrelevant variables

Model is overfitted

Q22. Why scale variables?

Ans. Needed when variables have different units

Improves model performance

Important for regularization

🔹 POLYNOMIAL REGRESSION

Q23. What is polynomial regression?

Ans. It models non-linear relationships using polynomial terms.

Q24. Difference between linear & polynomial regression

Ans. Linear Polynomial Straight line Curved line Degree = 1 Degree ≥ 2

Q25. When is polynomial regression used?

Ans. When data shows a curved pattern

Linear model fails

Q26. General equation

Ans.

Y

a X n + b X n − 1 + ⋯ + c $Y=aX^n+bX^{n-1}+\cdots+c$

Q27. Can polynomial regression be applied to multiple variables?

Ans. Yes It can include multiple variables with polynomial terms

Q28. Limitations of polynomial regression

Ans. Overfitting

Poor extrapolation

Sensitive to outliers

Q29. Methods to choose polynomial degree Ans. Cross-validation

Adjusted R^2

AIC / BIC

Error metrics (MSE, RMSE)

Q30. Importance of visualization

Ans. Helps identify non-linear patterns

Helps choose correct degree

Avoids overfitting

Q31. Polynomial regression in Python

Ans. from sklearn.preprocessing import PolynomialFeatures from sklearn.linear_model import LinearRegression

poly = PolynomialFeatures(degree=2) X_poly = poly.fit_transform(X)

model = LinearRegression() model.fit(X_poly, y)

Double-click (or enter) to edit

Double-click (or enter) to edit