-:Regression:-

1. What is Simple Linear Regression?
Ans- Simple Linear Regression is a statistical technique used to model the relationship between one independent variable (X) and one dependent variable (Y). It assumes a linear relationship and is represented by the equation **Y = mX + c**, where *m* is the slope and *c* is the intercept. It is mainly used for prediction and understanding the strength of relationships.

2. What are the key assumptions of Simple Linear Regression?
Ans- The key assumptions are:

- Linearity between X and Y
- Independence of observations
- Homoscedasticity (constant variance of errors)
- Normal distribution of residuals
- No significant outliers

Violation of these assumptions can lead to unreliable results.

3. What does the coefficient m represent in the equation Y=mX+c?

Ans- The coefficient **m** represents the **slope** of the regression line. It indicates the change in the dependent variable (Y) for a one-unit change in the independent variable (X). A positive value shows an increasing relationship, while a negative value shows a decreasing relationship.

4. What does the intercept c represent in the equation Y=mX+c?

Ans- The intercept **c** represents the value of Y when X equals zero. It provides a baseline value for the dependent variable. In some real-world cases, this value may not be meaningful, but it helps position the regression line correctly.

5. How do we calculate the slope m in Simple Linear Regression?

Ans- The slope **m** is calculated using the formula:

m=∑(X−Xˉ)(Y−Yˉ)∑(X−Xˉ)2m = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}m=∑(X−Xˉ)2∑(X−Xˉ)(Y−Yˉ)

It measures how strongly X and Y vary together relative to the variation in X.

6. What is the purpose of the least squares method in Simple Linear Regression?

Ans- The least squares method is used to find the best-fitting regression line by minimizing the sum of the squared differences between observed values and predicted values. This ensures that the model has the smallest possible overall prediction error.

7. How is the coefficient of determination (R²) interpreted in Simple Linear Regression?

Ans- R² measures the proportion of variance in the dependent variable explained by the independent variable. Its value ranges from 0 to 1. A higher R² indicates a better fit of the model to the data.

8. What is Multiple Linear Regression?

Ans- Multiple Linear Regression is an extension of Simple Linear Regression where one dependent variable is predicted using two or more independent variables. The general form is:

Y=b0+b1X1+b2X2+⋯+bnXnY = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_nY=b0+b1X1+b2X2+⋯+bnXn

9. What is the main difference between Simple and Multiple Linear Regression?

Ans- Simple Linear Regression uses only one independent variable, while Multiple Linear Regression uses two or more independent variables. Multiple regression captures more complex relationships but is also more prone to issues like multicollinearity.

10. What are the key assumptions of Multiple Linear Regression?

Asn- The assumptions include:

- Linearity.
- Independence of errors.
- Homoscedasticity
  Normality of residuals.
- No multicollinearity among predictors
  No autocorrelation.
  These ensure reliable coefficient estimates.

11. What is heteroscedasticity, and how does it affect the results of a Multiple Linear Regression model?

Ans- Heteroscedasticity occurs when the variance of residuals is not constant across all levels of predictors. It leads to inefficient estimates and unreliable hypothesis testing, making confidence intervals and p-values inaccurate.

12. How can you improve a Multiple Linear Regression model with high multicollinearity?

Ans- Multicollinearity can be reduced by:

- Removing highly correlated variables.
- Combining variables.
- Using Principal Component Analysis (PCA).

- Applying Ridge or Lasso regression.
  These methods stabilize coefficient estimates.

13. What are some common techniques for transforming categorical variables for use in regression models?

Ans- Common techniques include:

- Label Encoding
- One-Hot Encoding
- Dummy Variable Encoding
  One-Hot Encoding is most widely used as it avoids ordinal assumptions.

14. What is the role of interaction terms in Multiple Linear Regression?

Ans- Interaction terms capture the combined effect of two or more variables on the dependent variable. They show how the relationship between one predictor and the outcome changes depending on another predictor.

15. How can the interpretation of intercept differ between Simple and Multiple Linear Regression?

Ans- in Simple Linear Regression, the intercept is the expected value of Y when X = 0. In Multiple Linear Regression, it represents the expected value of Y when all independent variables are zero, which may not always be meaningful.

16. What is the significance of the slope in regression analysis, and how does it affect predictions?

Ans- The slope indicates the direction and magnitude of the relationship between predictors and the outcome. It directly influences predictions by determining how much Y changes with a change in X.

17. How does the intercept in a regression model provide context for the relationship between variables?

Ans- The intercept provides a reference point for predictions. It helps understand the baseline level of the dependent variable and positions the regression line relative to the data.

18. What are the limitations of using R^2 as a sole measure of model performance?

Ans- R^2 does not indicate causality or model correctness. It always increases with more variables, even irrelevant ones. It also does not reflect overfitting or prediction accuracy on new data.

19. How would you interpret a large standard error for a regression coefficient?

Ans- A large standard error indicates high uncertainty in the coefficient estimate. It suggests that the predictor may not be statistically significant or that multicollinearity or insufficient data exists.

20. How can heteroscedasticity be identified in residual plots, and why is it important to address it?

Ans- Heteroscedasticity is identified when residuals show a funnel or uneven spread pattern. Addressing it is important because it affects the reliability of statistical tests and confidence intervals.

21. What does it mean if a Multiple Linear Regression model has a high R^2 but low adjusted R^2?

Ans- This indicates that some predictors do not contribute meaningfully to the model. Adjusted R^2 penalizes unnecessary variables, so a low value suggests overfitting.

22. Why is it important to scale variables in Multiple Linear Regression?

Asn- Scaling ensures that variables with large magnitudes do not dominate the model. It improves numerical stability and is essential when using regularization techniques like Ridge or Lasso.

23. What is polynomial regression?

Ans- Polynomial regression is a form of regression where the relationship between X and Y is modeled as a polynomial equation. It is used when data shows a curved, non-linear pattern.

24. How does polynomial regression differ from linear regression?

Ans- Linear regression models straight-line relationships, while polynomial regression models curved relationships by including higher-degree terms. Despite this, polynomial regression is still linear in parameters.

25. When is polynomial regression used?

Ans- Polynomial regression is used when the relationship between variables is non-linear but still smooth and continuous. It helps capture curvature in data trends.

26. What is the general equation for polynomial regression?

Asn- The general equation is:

Y=b0+b1X+b2X2+⋯+bnXnY = b_0 + b_1X + b_2X^2 + \dots + b_nX^nY=b0+b1X+b2X2+⋯+bnXn

where *n* is the degree of the polynomial.

27. Can polynomial regression be applied to multiple variables?

Ans- Yes, polynomial regression can be applied to multiple variables by including polynomial terms and interaction terms for each predictor.

28. What are the limitations of polynomial regression?

Ans- Limitations include overfitting, poor extrapolation outside the data range, sensitivity to outliers, and difficulty in interpretation for higher-degree polynomials.

29. What methods can be used to evaluate model fit when selecting the degree of a polynomial?

Ans- Methods include:

- Cross-validation
- Adjusted R^2
- Mean Squared Error (MSE)
- AIC/BIC criteria

  These help balance bias and variance

30. Why is visualization important in polynomial regression?

Asn- Visualization helps understand the curvature of the data, detect overfitting or underfitting, and compare different polynomial degrees effectively.

31. How is polynomial regression implemented in Python?

Ans- Polynomial regression in Python is implemented using **PolynomialFeatures** from `sklearn.preprocessing` along with **LinearRegression**. The features are transformed into polynomial terms before fitting the model.