1.What is Simple Linear Regression?

Answer Simple linear regression is used to find the relationship between two variables by fitting the

best straight line through the data points

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

Answer **Key assumptions of Simple Linear Regression are:**

1. **Linearity**: The relationship between the independent and dependent variables is linear.
2. **Independence**: The residuals (errors) are independent.
3. **Homoscedasticity**: The variance of the residuals is constant across all levels of the independent variable.
4. **Normality**: The residuals are normally distributed.

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

Answer In the linear equation Y=mX+c, the coefficient m represents the slope of the line.

* The m tells us how steep the line is.
* If m is positive, the line goes up as we move to the right.
* If m is negative, the line goes down as we move to the right.
* Basically, m shows how much Y changes when X changes.

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

Answer In the equation Y=mX+cY = mX + c, the intercept cc represents the point where the line crosses the Y-axis.

* This happens when X=0X = 0.
* So, cc is the value of YY when XX is 0

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

Answer To calculate the slope m in simple linear regression, we can use the following formula:

m=∑(xi−xˉ)(yi−yˉ) / ∑(xi−xˉ)2

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

Answer >>The purpose of the least squares method in simple linear regression is to find the best-fitting line through a set of data points by minimizing the sum of the squared differences (residuals) between the observed values and the values predicted by the linear model. This method ensures that the line is as close as possible to all the data points, providing the most accurate representation of the relationship between the variables.

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

Answer>> How is the coefficient of determination (R²) interpreted in Simple Linear Regression

The coefficient of determination, denoted as R2R^2, is a measure of how well the regression line explains the variation in the dependent variable. It ranges from 0 to 1 and is interpreted as follows:

* R2=1R^2 = 1: The regression line perfectly fits the data, explaining all the variability in the dependent variable.
* 0<R2<10 < R^2 < 1: The regression line explains some, but not all, of the variability in the dependent variable. The closer R2R^2 is to 1, the better the model fits the data.
* R2=0R^2 = 0: The regression line does not explain any of the variability in the dependent variable.In essence, R2R^2 tells us the proportion of the variance in the dependenvariable that is predictable from the independent variable.

8- What is Multiple Linear Regression?

Answer **Multiple Linear Regression** is a statistical technique used to predict the value of a dependent variable based on multiple independent variables. Essentially, it helps us understand how different factors together influence an outcome.

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

Answer>> The main difference between simple and multiple linear regression is that simple linear regression uses only one independent variable to predict the target variable, while multiple linear regression uses two or more independent variables for prediction.

10 What are the key assumptions of Multiple Linear Regression?

Answer The key assumptions of Multiple Linear Regression:

1. **Linearity**: The relationship between the variables is a straight line.
2. **Independence**: Each data point is independent of the others.
3. **Equal Variance**: The spread of the data is the same for all values of the independent variables.
4. **Normality**: The data follows a normal distribution.
5. **No Correlation**: The independent variables are not too closely related to each other.

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

Answer **Heteroscedasticity** occurs when the variance of the residuals (errors) in a regression model is not constant across all levels of the independent variables. In other words, the spread of the residuals changes as the value of the independent variables changes.

**Impact on Multiple Linear Regression:**

* **Inaccurate Standard Errors**: Heteroscedasticity can lead to incorrect estimates of the standard errors of the regression coefficients, which affects the reliability of hypothesis tests and confidence intervals.
* **Biased Estimates**: The regression coefficients may still be unbiased, but the efficiency of the estimates is reduced, meaning they are less precise.
* **Invalid Statistical Inferences**: The presence of heteroscedasticity can invalidate the results of statistical tests, leading to incorrect conclusions about the significance of the independent variables.

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

Answer Apply methods like Ridge Regression or Lasso Regression, Elastic Net we can handle multicollinearity by adding a penalty to the regression coefficients.

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

Answer>>Here are some common techniques for transforming categorical variables for use in regression models:

1. **One-Hot Encoding**: Converts categorical features into binary vectors, with one column for each category.
2. **Label Encoding**: Assigns a unique integer to each category, but can introduce an artificial order.
3. **Ordinal Encoding**: Similar to Label Encoding but preserves the order of categories if they have a meaningful sequence.

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

Answer **Interaction terms** show how two or more variables together affect the outcome in a regression model.

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

Answer In **Simple Linear Regression**, the intercept represents the expected value of the dependent variable when the independent variable is zero.

In **Multiple Linear Regression**, the intercept represents the expected value of the dependent variable when all the independent variables are set to zero.

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

Answer>>The **slope** in regression analysis shows how much the dependent variable changes when the independent variable increases by one unit. It helps predict the outcome based on the input.

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

Answer The **intercept** in a regression model is the starting value of the dependent variable when all independent variables are zero. It gives a baseline to understand the relationship between variables.

18 What are the limitations of using R² as a sole measure of model performance?

Answer Limitations of using R² as a sole measure of model per:

1. **Overfitting**: High R² might mean the model is too complex.
2. **Scale Sensitivity**: R² can be misleading when comparing different scales.
3. **No Penalty for Complexity**: Adding more variables can inflate R².
4. **Limited Applicability**: R² is mainly for linear regression and not other models.

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

Answer A large standard error for a regression coefficient means the estimate is uncertain and could vary significantly with new data. It shows we're less sure about the true value of that coefficient.

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

Answer **Heteroscedasticity** can be identified in residual plots by looking for patterns where the spread of residuals (errors) changes across different levels of the independent variable. Specifically, if the residuals form a funnel shape, with the spread increasing or decreasing as the predicted values increase, it indicates heteroscedasticity.

**It is important to address for following reasons**:

* **Inaccurate Standard Errors**: Heteroscedasticity can lead to incorrect estimates of the standard errors of the regression coefficients, affecting hypothesis tests and confidence intervals.
* **Biased Estimates**: While the regression coefficients themselves may still be unbiased, their efficiency is reduced, making them less reliable.
* **Invalid Statistical Inferences**: The presence of heteroscedasticity can invalidate the results of statistical tests, leading to incorrect conclusions about the significance of the independent variables.

**21** What does it mean if a Multiple Linear Regression model has a high R² but low adjusted R²?

Answer )B- What does it mean if a Multiple Linear Regression model has a high R² but low adjusted R²

If a Multiple Linear Regression model has a high **R²** but a low **adjusted R²**, it suggests that:

1. **Overfitting**: The model may be overfitting the data, capturing noise rather than the true underlying relationship.
2. **Too Many Variables**: The high R² is likely due to including too many variables, some of which may not be truly significant.
3. **Adjusted R² Penalty**: Adjusted R² penalizes the addition of non-informative predictors, providing a more accurate measure of model performance.

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

Answer Scaling helps to standardize the data, reducing calculation time and improving the accuracy and efficiency of the model.

23. What is polynomial regression?

ANSWER>>> **Polynomial regression** is a way to fit a curve to data instead of a straight line. It helps capture more complex relationships between variables.

24 How does polynomial regression differ from linear regression?

ANSWER>> **Polynomial regression** differs from **linear regression** in that it fits a curve to the data, while linear regression fits a straight line. Polynomial regression can capture more complex relationships by including higher-degree terms (like x2x^2, x3x^3, etc.), whereas linear regression only includes the first-degree term (xx).

25 When is polynomial regression used?

ANSWER>> Polynomial regression is used when the relationship between the independent variable and the dependent variable is not linear, meaning it cannot be accurately captured by a straight line. It's helpful when the data shows a curved pattern, allowing for a better fit and more accurate predictions.

26 What is the general equation for polynomial regression?

ANSWER>> The general equation for polynomial regression is:

y=b0+b1x+b2x2+…+bnxn+ϵ

Where y is the dependent variable

* x is the independent variable
* b0,b1,…,bn are the coefficients
* ϵ is the error term

27 Can polynomial regression be applied to multiple variables?

Answer **Yes, polynomial regression can handle multiple variables**. It means we can include not just the original variables but also their squared and combined terms to create a more detailed model.

28 What are the limitations of polynomial regression ?

Limitations of Polynomial regressions are:

Answer  **Overfitting**: Higher-degree polynomials can fit the training data very well, capturing noise along with the underlying data pattern. This can lead to poor performance on new, unseen data.

 **Complexity**: As the degree of the polynomial increases, the model becomes more complex and harder to interpret.

 **Computation**: Higher-degree polynomials require more computational resources and can be slower to train and evaluate.

 **Extrapolation**: Polynomial regression can behave unpredictably outside the range of the training data, leading to unreliable predictions.

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

Answer We can use these method to evaluate model fit when selecting the degree of a polynomial?:

1. **Mean Squared Error (MSE)**: Lower is better.
2. **R-squared (R²)**: Higher is better.
3. **Cross-Validation**: Check if the model works well on different data sets.

30 Why is visualization important in polynomial regression?

Answer>> Visualization is important in polynomial regression because it helps us see how well the model fits the data and makes it easier to understand and explain the results

31 How is polynomial regression implemented in Python?

Answer>> Steps to implement polynomial regression in python are as follows:

1 **Import Libraries**:

import numpy as np

import matplotlib.pyplot as plt

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import PolynomialFeatures

from sklearn.linear\_model import LinearRegressionfrom sklearn.metrics import mean\_squared\_error, r2\_score

2 **Generate Data**:

np.random.seed(0)

x = 2 - 3 \* np.random.normal(0, 1, 20)

y = x - 2 \* (x \*\* 2) + 0.5 \* (x \*\* 3) + np.random.normal(-3, 3, 20)

3 **Split Data**:

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size=0.2, random\_state=0)

4 **Transform Data**:

poly = PolynomialFeatures(degree=2)

x\_poly = poly.fit\_transform(x\_train.reshape(-1, 1))

5 **Fit Model**:

model = LinearRegression()

model.fit(x\_poly, y\_train)

6 **Predict and Evaluate**:

y\_poly\_pred = model.predict(x\_poly)

rmse = np.sqrt(mean\_squared\_error(y\_train, y\_poly\_pred))

r2 = r2\_score(y\_train, y\_poly\_pred)

print('RMSE:', rmse)

print('R2:', r2)

7 **Plot Results**:

plt.scatter(x\_train, y\_train, color='red')

plt.plot(x\_train, y\_poly\_pred, color='blue')

plt.title('Polynomial Regression')

plt.xlabel('X')

plt.ylabel('Y')

plt.show()