26 MARCH

Q1. Explain the difference between simple linear regression and multiple linear regression. Provide an example of each.

Simple Linear Regression : In simple linear regression, there is a single independent variable (predictor) that is used to predict a continuous dependent variable. The relationship is modelled as a straight line (a linear equation). For example, you might use simple linear regression to predict a person's weight based on their height.

Multiple Linear Regression : In multiple linear regression, there are two or more independent variables used to predict a continuous dependent variable. The relationship is still linear, but it involves multiple predictors. For example, you might use multiple linear regression to predict a house's price based on factors like square footage, number of bedrooms, and location.

Q2. Discuss the assumptions of linear regression. How can you check whether these assumptions hold in a given dataset?

The common assumptions of linear regression include linearity, independence of errors, constant variance (homoscedasticity), and normally distributed errors. You can check these assumptions through diagnostic plots (e.g., residual plots), statistical tests, and by examining the data visually.

Q3. How do you interpret the slope and intercept in a linear regression model? Provide an example using a real world scenario.

Slope (Coefficient) : The slope represents the change in the dependent variable for a one unit change in the independent variable, holding all other variables constant. For example, in a simple linear regression predicting salary (dependent variable) based on years of experience (independent variable), a slope of 2,000 would mean that for each additional year of experience, the salary is expected to increase by $2,000.

Intercept : The intercept is the predicted value of the dependent variable when all independent variables are set to zero. In the salary example, it could represent the starting salary for someone with zero years of experience.

Q4. Explain the concept of gradient descent. How is it used in machine learning?

Gradient descent is an optimization algorithm used in machine learning to minimize the loss or cost function of a model. It works by iteratively adjusting model parameters in the direction of steepest descent (the negative gradient) to find the minimum of the cost function.

Gradient descent is used to train various machine learning models, including linear regression, neural networks, and more.

Q5. Describe the multiple linear regression model. How does it differ from simple linear regression?

Multiple linear regression is an extension of simple linear regression where there are two or more independent variables used to predict a continuous dependent variable. It differs from simple linear regression in that it accounts for the potential influence of multiple predictors on the dependent variable, allowing for a more complex and nuanced modeling of relationships.

Q6. Explain the concept of multicollinearity in multiple linear regression. How can you detect and address this issue?

Multicollinearity occurs when two or more independent variables in a multiple linear regression model are highly correlated with each other. It can lead to unstable coefficient estimates and make it challenging to interpret the individual contributions of predictors. You can detect multicollinearity using correlation matrices or variance inflation factors (VIFs) and address it by removing one of the correlated variables, combining them, or using regularization techniques like Ridge or Lasso regression.

Q7. Describe the polynomial regression model. How is it different from linear regression?

Polynomial regression is a form of regression where the relationship between the independent variable(s) and the dependent variable is modeled as an nth degree polynomial equation. Unlike linear regression, which assumes a linear relationship, polynomial regression can capture nonlinear relationships in the data.

Q8. What are the advantages and disadvantages of polynomial regression compared to linear regression? In what situations would you prefer to use polynomial regression?

Advantages:

- Can model complex, nonlinear relationships.
- Provides a more flexible fit to the data.
- Can capture interactions between variables.

Disadvantages:

- Susceptible to overfitting when using high degree polynomials.
- Interpretability can be challenging.
- Extrapolation can lead to inaccurate predictions.