Q1. **Difference between simple linear regression and multiple linear regression:**

- **Simple linear regression** involves predicting a dependent variable using only one independent variable. For example, predicting house prices based on the area of the house.
- **Multiple linear regression**, on the other hand, involves predicting a dependent variable using two or more independent variables. For example, predicting house prices based on the area, number of bedrooms, and location.

Q2. **Assumptions of linear regression and how to check them:**

- Assumptions include linearity, independence of errors, constant variance of errors (homoscedasticity), and normality of errors.
- These assumptions can be checked using diagnostic plots (e.g., residuals vs. fitted values plot), statistical tests (e.g., Shapiro-Wilk test for normality), and examining correlation matrices for multicollinearity.

Q3. **Interpretation of slope and intercept in linear regression:**

- The **slope** represents the change in the dependent variable for a one-unit change in the independent variable, holding other variables constant.
- The **intercept** represents the value of the dependent variable when all independent variables are zero.
- For example, in a regression model predicting exam scores based on study hours, the slope indicates the increase in exam score for each additional hour of study, and the intercept indicates the expected exam score when no study hours are logged.

Q4. **Concept of gradient descent and its use in machine learning:**

- Gradient descent is an optimization algorithm used to minimize the loss function in machine learning models.
- It works by iteratively adjusting the parameters of the model in the direction of the steepest descent of the loss function, until convergence is achieved.
- Gradient descent is used to update the coefficients in linear regression models to minimize the difference between predicted and actual values.

Q5. **Description of multiple linear regression model:**

- Multiple linear regression predicts a dependent variable using two or more independent variables.
- The model equation is of the form: $Y=\beta 0+\beta 1X1+\beta 2X2+...+\beta nXn+\epsilon Y=\beta 0+\beta 1X1+\beta 2X2+...+\beta nXn+\epsilon$, where YY is the dependent variable, $X1,X2,...,XnX1,X2,...,Xn$ are the independent variables, $\beta 0,\beta 1,\beta 2,...,\beta n\beta 0,\beta 1,\beta 2,...,\beta n$ are the coefficients, and $\epsilon\epsilon$ is the error term.

Q6. **Explanation of multicollinearity in multiple linear regression and its detection:**

- Multicollinearity occurs when independent variables in a multiple linear regression model are highly correlated with each other.
- It can be detected by examining correlation matrices or variance inflation factors (VIFs) for each independent variable.
- To address multicollinearity, variables can be removed, or techniques like principal component analysis (PCA) or ridge regression can be used.

Q7. **Description of polynomial regression model and its difference from linear regression:**

- Polynomial regression is a form of regression analysis in which the relationship between the independent variable XX and the dependent variable YY is modeled as an nth degree polynomial.
- Unlike linear regression, which assumes a linear relationship between XX and YY, polynomial regression can capture nonlinear relationships.

Q8. **Advantages and disadvantages of polynomial regression compared to linear regression:**

- Advantages: Can model nonlinear relationships, flexible, captures more complex patterns.
- Disadvantages: Can overfit the data, more complex interpretation, may perform poorly outside the range of observed data.
- Polynomial regression is preferred when there's evidence of nonlinear relationships between variables and when better predictive performance is desired despite the risk of overfitting.