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

**Simple Linear Regression** involves predicting a dependent variable (Y) based on a single independent variable (X). The relationship is modeled with a straight line, described by the equation:  
Y=β0+β1X+ϵY = \beta\_0 + \beta\_1X + \epsilon

**Example:**  
Predicting a person's weight (Y) based on their height (X).

**Multiple Linear Regression** extends simple linear regression by using multiple independent variables (X1, X2, ..., Xn) to predict the dependent variable. The model equation is:  
Y=β0+β1X1+β2X2+...+βnXn+ϵY = \beta\_0 + \beta\_1X\_1 + \beta\_2X\_2 + ... + \beta\_nX\_n + \epsilon

**Example:**  
Predicting a person's weight (Y) based on their height (X1), age (X2), and activity level (X3).

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

Linear regression relies on several assumptions:

1. **Linearity**: There is a linear relationship between the independent and dependent variables.
   * **Check**: Plot the data and observe if a straight line fits the points well. Residual plots can also help verify this.
2. **Independence**: The observations are independent of each other.
   * **Check**: Use the Durbin-Watson test for autocorrelation in the residuals.
3. **Homoscedasticity**: The variance of errors is constant across all levels of the independent variables.
   * **Check**: Plot residuals against predicted values. If the variance is not constant (e.g., a fan-shaped pattern), this assumption is violated.
4. **Normality of residuals**: The residuals (errors) should be normally distributed.
   * **Check**: Use Q-Q plots or a histogram of residuals. Additionally, the Shapiro-Wilk test can assess normality.
5. **No multicollinearity**: Independent variables should not be highly correlated with each other.
   * **Check**: Use Variance Inflation Factor (VIF) or correlation matrices.

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

* **Intercept (β0\beta\_0)**: The value of the dependent variable when all independent variables are zero. It represents the baseline or starting value.
* **Slope (β1\beta\_1)**: The change in the dependent variable for a one-unit change in the independent variable. It indicates the strength and direction of the relationship.

**Example:**  
In predicting salary (Y) based on years of experience (X), the equation might be:  
Salary=30,000+5,000×Years of Experience\text{Salary} = 30,000 + 5,000 \times \text{Years of Experience}

* The intercept (β0=30,000\beta\_0 = 30,000) means that with 0 years of experience, the salary is $30,000.
* The slope (β1=5,000\beta\_1 = 5,000) means that for every additional year of experience, the salary increases by $5,000.

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

**Gradient Descent** is an optimization algorithm used to minimize the cost function in machine learning models, particularly in linear regression. The algorithm iteratively adjusts the model parameters (like the slope and intercept) to reduce the error (loss) between the predicted and actual values.

1. Start with initial values for the parameters.
2. Compute the gradient (partial derivative) of the cost function.
3. Update the parameters in the direction of the negative gradient (to minimize the cost).
4. Repeat this process until the cost function converges to a minimum value.

In machine learning, gradient descent is used to optimize various models, ensuring they fit the data well by minimizing the loss function.

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

**Multiple Linear Regression** predicts a dependent variable based on more than one independent variable. It takes into account the simultaneous influence of multiple predictors on the target variable, making it more suitable for complex relationships.

The equation for multiple linear regression is:  
Y=β0+β1X1+β2X2+...+βnXn+ϵY = \beta\_0 + \beta\_1X\_1 + \beta\_2X\_2 + ... + \beta\_nX\_n + \epsilon

**Difference from Simple Linear Regression**:

* **Simple Linear Regression** has one independent variable, whereas **Multiple Linear Regression** uses two or more independent variables.
* Multiple linear regression can capture more complex relationships, while simple linear regression is suited for modeling linear relationships with one variable.

**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 are highly correlated with each other, making it difficult to estimate the individual effect of each variable on the dependent variable. This can lead to unstable coefficients and inflated standard errors.

**Detection:**

* **Correlation Matrix**: High correlations (close to 1 or -1) between independent variables indicate multicollinearity.
* **Variance Inflation Factor (VIF)**: A VIF greater than 10 indicates a high degree of multicollinearity.

**Addressing Multicollinearity:**

1. Remove one of the correlated variables.
2. Combine correlated variables through techniques like Principal Component Analysis (PCA).
3. Use regularization methods like Ridge or Lasso regression to penalize large coefficients.

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

**Polynomial Regression** is a type of regression where the relationship between the independent variable (X) and the dependent variable (Y) is modeled as an nth-degree polynomial rather than a straight line.

The equation for polynomial regression is:  
Y=β0+β1X+β2X2+...+βnXn+ϵY = \beta\_0 + \beta\_1X + \beta\_2X^2 + ... + \beta\_nX^n + \epsilon

**Difference from Linear Regression**:

* Linear regression assumes a straight-line relationship, whereas polynomial regression models more complex, nonlinear relationships.
* Polynomial regression can better fit curves in data, whereas linear regression is restricted to straight-line trends.

**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 of Polynomial Regression**:

1. Can capture nonlinear relationships between variables.
2. Provides a better fit for data that shows curvilinear patterns.

**Disadvantages of Polynomial Regression**:

1. Can lead to overfitting if the degree of the polynomial is too high.
2. Less interpretable compared to simple linear regression.
3. Requires careful selection of the degree of the polynomial to avoid overfitting or underfitting.

**When to use Polynomial Regression**:

* When there is evidence of a nonlinear relationship between the independent and dependent variables (e.g., growth trends, quadratic behavior).
* In situations where the data shows a curve that linear regression cannot fit well.