**Interview Questions:**

**1.What assumptions does linear regression make?**

Linear regression relies on several key assumptions:

* Linearity: The relationship between independent and dependent variables is linear.
* Independence: Observations are independent of each other.
* Homoscedasticity: Constant variance of residuals (errors).
* Normality: Residuals are normally distributed.
* No Multicollinearity: Independent variables are not highly correlated.

**2.How do you interpret the coefficients?**

* Each coefficient represents the expected change in the dependent variable for a one-unit change in the corresponding independent variable, keeping other variables constant.
  + Example: If the coefficient of X₁ is 2.5, it means that for every unit increase in X₁, the predicted value of Y increases by 2.5, assuming all other variables remain unchanged.

**3.What is R² score and its significance?**

* R² (Coefficient of Determination) measures how well the regression model fits the data.
* It represents the proportion of variance in the dependent variable that is explained by the independent variables.
  + Formula:
* Interpretation:
  + R² = 0 → Model explains none of the variability.
  + R² = 1 → Model explains all variability perfectly.
  + A higher R² indicates a better fit (but not always better generalization).

**4.When would you prefer MSE over MAE?**

* MSE (Mean Squared Error) penalizes larger errors more heavily due to squaring.
* Use MSE when:
  + You want to penalize large deviations more strongly.
  + You care about mathematical convenience (since it’s differentiable).
* Use MAE (Mean Absolute Error) when:
  + You want a robust metric less sensitive to outliers.

**5.How do you detect multicollinearity?**

**Multicollinearity** occurs when independent variables are highly correlated.  
Detection methods:

* Correlation Matrix: Check correlation values between independent variables (>0.8 indicates high correlation).
* Variance Inflation Factor (VIF):
  + If VIF > 5 or 10, it indicates multicollinearity.

**6.What is the difference between simple and multiple regression?**

| **Type** | **Description** | **Example** |
| --- | --- | --- |
| Simple Linear Regression | Uses one independent variable to predict the dependent variable. | Y = β₀ + β₁X + ε |
| Multiple Linear Regression | Uses two or more independent variables. | Y = β₀ + β₁X₁ + β₂X₂ + ... + ε |

**7.Can linear regression be used for classification?**

* No, linear regression is designed for continuous dependent variables.
* However, a variant called Logistic Regression is used for classification tasks, which models the probability of class membership.

**8.What happens if you violate regression assumptions?**

Violating assumptions can lead to:

* Non-linearity: Poor model fit and biased coefficients.
* Heteroscedasticity: Unreliable significance tests (p-values).
* Non-normal errors: Invalid confidence intervals.
* Multicollinearity: Inflated standard errors and unstable coefficients.
* Autocorrelation: Errors are correlated, violating independence, leading to biased estimates.