1. What assumptions does linear regression make?

- Linearity: Relationship between features and target is linear

- Independence: Observations are independent of each other

- Homoscedasticity: Residuals have constant variance

- Normality: Residuals are normally distributed

- No multicollinearity: Features aren't highly correlated with each other

- No auto-correlation: Residuals aren't correlated with each other

2. How do you interpret the coefficients?

- Coefficients represent the change in the target variable for a one-unit change in the feature, holding other features constant.

- Positive coefficient means target increases with the feature.

- Negative coefficient means target decreases with the feature.

3. What is R² score and its significance?

- R² (R-squared) measures the proportion of variance in the target explained by the features.

- Range: 0 to 1 (higher is better).

- Represents goodness of fit, but can be misleading with many features.

4. When would you prefer MSE over MAE?

- MSE is more sensitive to outliers (squares errors), so use when large errors are particularly bad.

- MAE gives equal weight to all errors, use when all errors are equally important.

5. How do you detect multicollinearity?

- Calculate Variance Inflation Factor (VIF) - VIF > 5-10 indicates multicollinearity.

- Examine correlation matrix for highly correlated features (>0.8 or <-0.8).

- Check for unstable coefficient estimates when adding/removing features.

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

- Simple regression uses one feature to predict the target.

- Multiple regression uses two or more features to predict the target.

7. Can linear regression be used for classification?

- Technically yes (logistic regression is a type of linear regression for classification).

- But standard linear regression isn't ideal for classification as it predicts continuous values.

8. What happens if you violate regression assumptions?

- Coefficient estimates may be biased or inefficient.

- P-values and confidence intervals may be inaccurate.

- Predictions may be less reliable.

- Model may not generalize well to new data.