

Task 3 - Linear Regression Interview Q&A

1. What assumptions does linear regression make?

1. Linearity: relationship between inputs and target is linear
2. No multicollinearity: features should not be highly correlated
3. Homoscedasticity: constant variance of residuals
4. Normality of residuals: residuals should follow a normal distribution
5. Independence of errors

2. How do you interpret the coefficients?

Each coefficient shows how much the target variable changes with a 1-unit increase in that feature, holding others constant.

Example: $\text{price} = 3000 \times \text{area} + 10000 \times \text{bedrooms}$ means 1 sqft increase adds Rs.3000 to price.

3. What is R^2 score and its significance?

R^2 tells how much of the variance in the target is explained by the model.

$R^2 = 1$ means perfect fit; $R^2 = 0$ means the model does no better than predicting the average.

4. When would you prefer MSE over MAE?

MSE penalizes larger errors more, so it's useful when large errors are worse (e.g., price prediction).

MAE treats all errors equally and is better when you want a more general error measure.

5. How do you detect multicollinearity?

Use a correlation matrix or Variance Inflation Factor (VIF).

$VIF > 5$ (or 10) suggests multicollinearity between features.

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

Simple regression uses one input feature; multiple regression uses two or more.

Example: $\text{price} = a \times \text{area}$ (simple), vs $\text{price} = a \times \text{area} + b \times \text{bedrooms}$ (multiple).

7. Can linear regression be used for classification?

Not recommended. Linear regression is for continuous outputs.

Use Logistic Regression for classification tasks.

8. What happens if you violate regression assumptions?

Violating assumptions can lead to:

- Biased or inaccurate predictions
- Misleading coefficients
- Reduced model performance

Patterns in residuals indicate missed trends.