

Linear Regression Interview – Polished Questions & Answers

1. What is Linear Regression? Linear regression is a supervised learning algorithm used to predict a continuous target variable by modeling the linear relationship between input features and the output. Mathematically, it is represented as $y = w_0 + w_1x_1 + \dots + w_nx_n + \epsilon$.
2. What cost function is used and why? Linear regression uses Mean Squared Error (MSE), which is the average of the squared differences between actual and predicted values. It is convex, differentiable, and penalizes large errors.
3. How are parameters learned? Parameters are learned using Gradient Descent, which iteratively updates weights in the opposite direction of the gradient of the loss function to minimize error.
4. Assumptions of Linear Regression - Linearity - Independence of errors - Homoscedasticity - Normality of errors - No multicollinearity
5. What is overfitting and how do you prevent it? Overfitting occurs when the model learns noise instead of signal. It can be prevented using regularization (Ridge/Lasso), cross-validation, feature selection, and more data.
6. Ridge vs Lasso Ridge uses L2 penalty and shrinks coefficients but keeps all features. Lasso uses L1 penalty and sets some coefficients to zero, performing feature selection.
7. Multicollinearity Occurs when features are highly correlated. It makes coefficients unstable and hard to interpret. Handled by removing features, using Ridge, or PCA.
8. Bias-Variance Tradeoff Bias is error from oversimplification. Variance is error from sensitivity to data. Regularization increases bias slightly but reduces variance.
9. Feature Scaling Scaling speeds up gradient descent convergence and ensures equal feature contribution. Common methods: Z-score standardization and Min-Max scaling.
10. When not to use Linear Regression - Classification problems - Clustering tasks - Non-linear relationships