Q1. What is Lasso Regression, and how does it differ from other regression techniques?

Lasso Regression (Least Absolute Shrinkage and Selection Operator) is a linear regression technique that adds a **penalty on the absolute values of the coefficients** (L_1 norm) to the ordinary least squares (OLS) cost function. It both **shrinks coefficients** and can **set some coefficients exactly to zero**, effectively performing feature selection.

- **Lasso Cost Function**:

$$\text{RSS}_{lasso} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Where:

- $\lambda \ge 0$ is the regularization parameter.
- β_j are the regression coefficients.

Differences from other regression techniques:

Regression Type	Penalty	Feature Selection	Coefficient Shrinkage
OLS	None	No	No
Ridge	L_2	No	Yes (toward 0)
Lasso	L_1	Yes (can be 0)	Yes

Q2. What is the main advantage of using Lasso Regression in feature selection?

- Lasso can **shrink some coefficients exactly to zero**, effectively **removing irrelevant or less important features** from the model.
- This helps with:

1. **Reducing model complexity**
2. **Improving interpretability**
3. **Avoiding overfitting** when many predictors are present

Ridge cannot do this—it only shrinks coefficients toward zero but never to zero.

Q3. How do you interpret the coefficients of a Lasso Regression model?

1. **Non-zero coefficients** indicate features that **contribute to predicting the response**.
2. **Magnitude of coefficients** reflects the relative importance (after standardization).
3. **Zero coefficients** mean the feature is **effectively excluded** from the model.
4. Coefficients are **biased** due to regularization, so interpret them cautiously in a causal sense.

Important: Always **standardize variables** before comparing coefficients because Lasso penalizes coefficients equally.

Q4. What are the tuning parameters that can be adjusted in Lasso Regression, and how do they affect the model's performance?

The primary tuning parameter is:

1. **λ\lambdaλ (regularization strength)**:
    - **Small λ≈0\lambda \approx 0λ≈0** → Lasso behaves like OLS (less shrinkage, more variance).
    - **Large λ\lambdaλ** → More coefficients shrink to zero (simpler model, more bias).

Other considerations (less common):

2. **Max iterations** (for solver convergence)

3. **Tolerance** (for convergence threshold)

Effect on performance:

- Choosing λ\\lambdaλ balances the **bias-variance tradeoff**.
- Optimal λ\\lambdaλ gives **good prediction accuracy** and **automatic feature selection**.

Q5. Can Lasso Regression be used for non-linear regression problems? If yes, how?

Yes, indirectly:

1. **Transform non-linear relationships into linear form** using:
   - Polynomial features
   - Interaction terms
   - Basis expansions (e.g., splines)
2. Apply **Lasso** to the expanded set of features.

- Lasso can then select **important non-linear terms**.
- However, Lasso itself **does not inherently model non-linearity**; you need to engineer features first.

Q6. What is the difference between Ridge Regression and Lasso Regression?

Feature	Ridge Regression	Lasso Regression
Penalty Type	L2L_2L2 (squared coefficients)	L1L_1L1 (absolute coefficients)
Coefficient Shrinkage	Yes, toward zero	Yes, toward zero

Feature Selection	No (all coefficients remain)	Yes (some coefficients can be zero)
Bias-Variance Tradeoff	Reduces variance, adds bias	Can reduce variance, add bias, and simplify model
Use case	Multicollinearity, small shrinkage	Feature selection, sparse models

---

Q7. Can Lasso Regression handle multicollinearity in the input features? If yes, how?

- **Partially**: Lasso can handle correlated features by **selecting one feature and shrinking others to zero**.
- However:
    - If features are **highly correlated**, Lasso may arbitrarily pick one feature and drop the rest.
    - Ridge is better if you want to **retain all correlated features** but stabilize coefficients.

So, Lasso helps **reduce dimensionality** in multicollinear settings but may not preserve all correlated predictors.

Q8. How do you choose the optimal value of the regularization parameter (λ\\lambdaλ) in Lasso Regression?

- **Cross-Validation (CV)**: Most common method
    - Perform **k-fold CV** for a range of λ\\lambdaλ values
    - Choose λ\\lambdaλ that **minimizes the mean squared error (MSE)** on validation sets
- **Information criteria (less common)**: AIC, BIC
- **Plotting techniques**: Lasso path plots show how coefficients shrink as λ\\lambdaλ increases

✅ **Rule of Thumb:**

- **Small λ\lambdaλ** → more features, less bias, more variance
- **Large λ\lambdaλ** → fewer features, more bias, less variance

---