# Ridge, Lasso & Elastic Net Regression

*Regularization techniques to prevent overfitting and improve model generalization*

📄 Download as PDF

## Ridge Regression (L2 Regularization)

**Cost Function:** RSS + $\lambda\sum\beta^2_j$

### Key Characteristics:

- Adds the **squared magnitude** of coefficients as penalty
- **Shrinks coefficients** towards zero but never exactly to zero
- All features are retained in the model
- Works well when many features have small-to-medium effects

### Interview Points:

- **Never eliminates features** - keeps all predictors
- Better when you believe all features contribute to the outcome
- More stable when features are highly correlated (multicollinearity)
- The penalty term helps with numerical stability in matrix inversion
- $\lambda$ (lambda) controls regularization strength: higher $\lambda$ = more shrinkage

## Lasso Regression (L1 Regularization)

**Cost Function:** RSS + $\lambda\sum|\beta_j|$

### Key Characteristics:

- Adds the **absolute value** of coefficients as penalty
- Can shrink coefficients **exactly to zero**
- Performs **automatic feature selection**
- Creates sparse models

### Interview Points:

- **Built-in feature selection** - eliminates irrelevant features
- Useful when you suspect only a subset of features are important
- Interpretable models due to sparsity
- Can struggle when features are highly correlated (may randomly select one)
- Not uniquely defined when features are correlated

## Elastic Net

`Cost Function:` $RSS + \lambda_1\sum|\beta_j| + \lambda_2\sum\beta^2_j$

**Key Characteristics:**

- **Combines L1 and L2 penalties**
- Balances between Ridge and Lasso
- Controlled by mixing parameter $\alpha$ (alpha)
- Gets benefits of both methods

**Interview Points:**

- **Best of both worlds** - feature selection + handles multicollinearity
- Preferred when features are correlated AND you want feature selection
- More robust than Lasso in high-dimensional settings
- $\alpha$ parameter: $\alpha=1$ is pure Lasso, $\alpha=0$ is pure Ridge, $0<\alpha<1$ is Elastic Net
- Generally performs better than Lasso when multiple features are correlated

## Critical Interview Comparisons

| Aspect | Ridge | Lasso | Elastic Net |
|---|---|---|---|
| **Feature Selection** | No | Yes | Yes |
| **Multicollinearity** | Handles well | Struggles | Handles well |
| **Sparsity** | No | Yes | Yes |
| **Computational** | Closed-form solution | Iterative | Iterative |

## When to Use What?

- **Ridge:** Many relevant features, multicollinearity present, want to keep all features

- **Lasso:** Need interpretability, suspect many irrelevant features, want automatic selection
- **Elastic Net:** High-dimensional data with correlated features, want feature selection with stability

## Common Interview Questions

**Q: Why does Lasso give exact zeros but Ridge doesn't?**
**A:** Due to the geometry of constraints - L1 constraint (diamond shape) has corners where coefficients can be exactly zero, while L2 constraint (circular) doesn't have sharp corners.

**Q: How do you choose λ?**
**A:** Use cross-validation (GridSearchCV or RandomizedSearchCV). Plot cross-validation error vs. λ to find optimal value.

**Q: What's the bias-variance tradeoff?**
**A:** Regularization increases bias but reduces variance, leading to better generalization. Higher λ = more bias, less variance.

**Q: Can you use these for classification?**
**A:** Yes! Logistic regression with L1/L2/Elastic Net regularization is common for classification tasks.

**Key Takeaway:** Elastic Net is often the safest default choice in practice, as it combines the strengths of both Ridge and Lasso while mitigating their individual weaknesses.