Comprehensive Guide to Regression: From Basic Slopes to Lasso Shrinkage

Instructor: Dr. Smith
Course: Advanced Data Science

August 6, 2025

1 Understanding Slope in Linear Regression

1.1 Geometric Interpretation

The slope (β_1) in the equation $Y = \beta_0 + \beta_1 X$ represents:

- **Steepness**: Rate of change in Y per unit change in X
- **Direction**: Positive or negative relationship
- **Angle**: Tangent of the angle between the regression line and x-axis

1.2 Step-by-Step Calculation

[Test Scores vs Study Hours] Consider this dataset:

Hours Studied (X)	Test Score (Y)
1	50
2	60
3	70
4	80

Step 1: Calculate Means

$$\bar{X} = \frac{1+2+3+4}{4} = 2.5$$
$$\bar{Y} = \frac{50+60+70+80}{4} = 65$$

Step 2: Compute Covariance

$$\begin{aligned}\text{Cov}(X, Y) &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n} \\ &= \frac{(1 - 2.5)(50 - 65) + \cdots + (4 - 2.5)(80 - 65)}{4} \\ &= \frac{22.5 + 2.5 + 2.5 + 22.5}{4} = 12.5\end{aligned}$$

Step 3: Compute Variance of X

$$\begin{aligned}\text{Var}(X) &= \frac{\sum (X_i - \bar{X})^2}{n} \\ &= \frac{(-1.5)^2 + (-0.5)^2 + (0.5)^2 + (1.5)^2}{4} \\ &= \frac{2.25 + 0.25 + 0.25 + 2.25}{4} = 1.25\end{aligned}$$

Step 4: Calculate Slope (β_1)

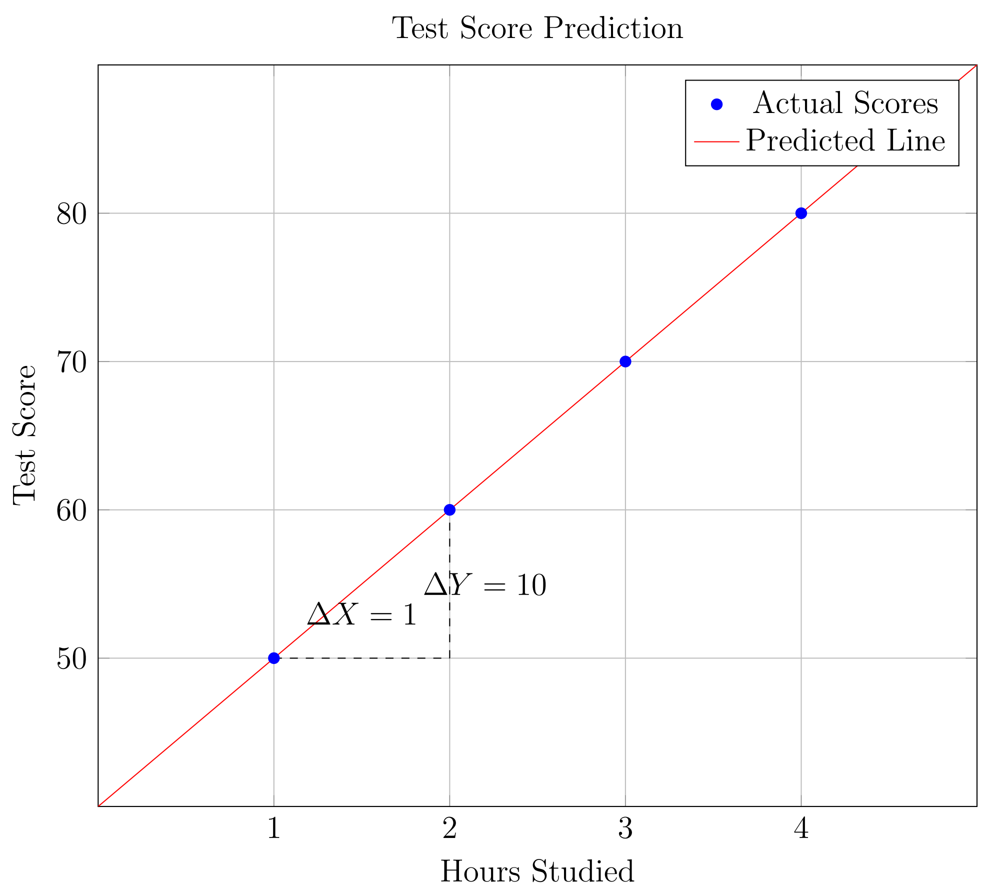
$$\beta_1 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{12.5}{1.25} = 10$$

Step 5: Calculate Intercept (β_0)

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} = 65 - 10 \times 2.5 = 40$$

Final Equation:

$$\hat{Y} = 40 + 10X$$



2 Lasso Regression: Slope Shrinkage

2.1 The Shrinkage Process

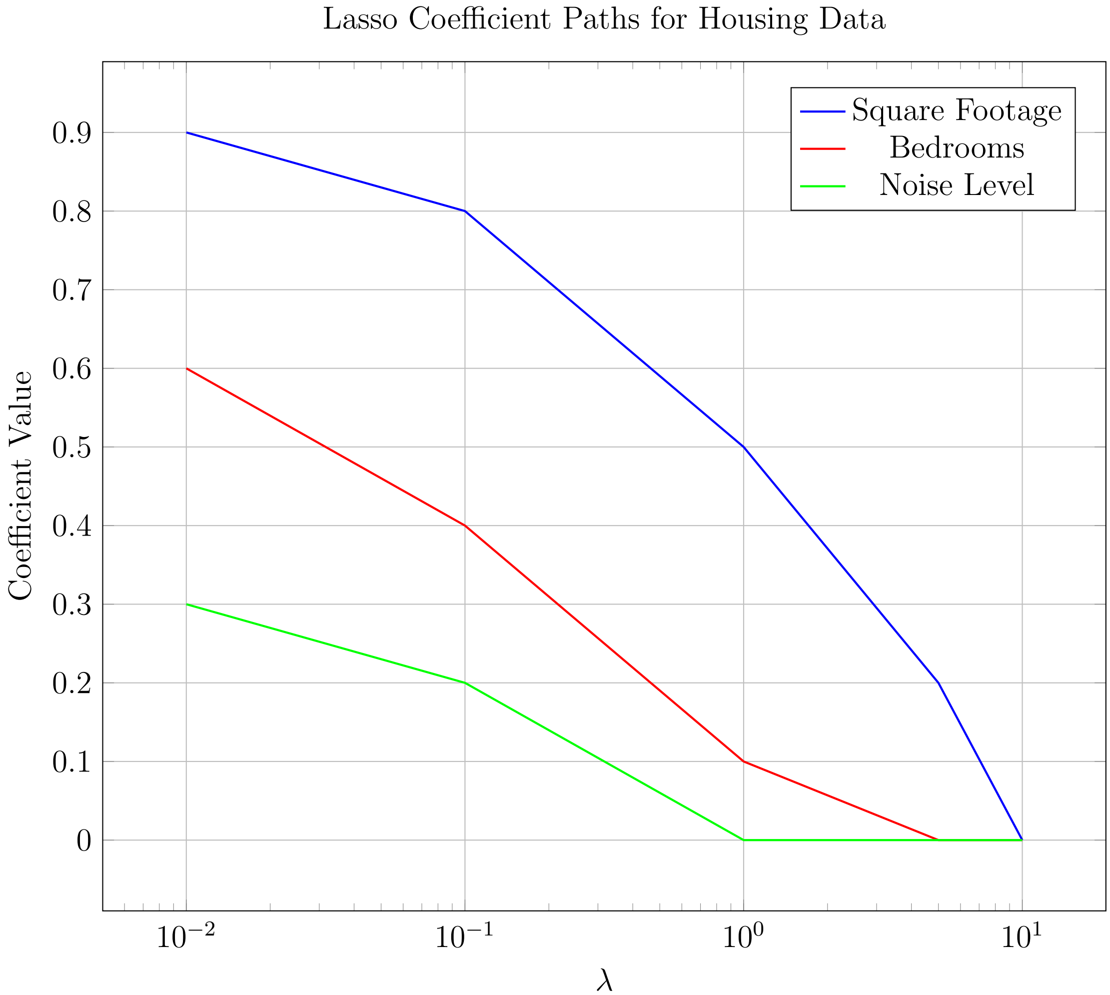
Lasso regression modifies the slope calculation by adding a penalty term:

$$\text{Minimize: } \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2 + \lambda|\beta_1|$$

Key Effects of λ

- As λ increases, $|\beta_1|$ is forced to decrease
- At some $\lambda_{critical}$, β_1 becomes exactly zero
- Different features have different $\lambda_{critical}$ values

2.2 Visualizing Shrinkage Paths



2.3 Why Shrinkage Matters

Before Lasso	**After Lasso**
Price = 50+20(SqFt)+10(Bed)−5(Noise)	Price = 48+18(SqFt)+8(Bed)+0(Noise)
• All features included • May overfit to noise • Harder to interpret	• Only important features remain • Reduces overfitting • Clearer interpretation

3 Complete Mathematical Derivation

3.1 Normal Equations for Linear Regression

Deriving the slope by minimizing SSE:

$$\begin{aligned} \frac{\partial}{\partial \beta_0} \sum (Y_i - \beta_0 - \beta_1 X_i)^2 = 0 \\ -2 \sum (Y_i - \beta_0 - \beta_1 X_i) = 0 \\ n\beta_0 = \sum Y_i - \beta_1 \sum X_i \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \beta_1} \sum (Y_i - \beta_0 - \beta_1 X_i)^2 = 0 \\ -2 \sum X_i (Y_i - \beta_0 - \beta_1 X_i) = 0 \\ \sum X_i Y_i = \beta_0 \sum X_i + \beta_1 \sum X_i^2 \end{aligned}$$

3.2 Lasso Modification

The Lasso objective function is non-differentiable at $\beta_j = 0$. We use subgradient conditions:

$$\frac{\partial}{\partial \beta_j} (\text{SSE} + \lambda |\beta_j|) = 0$$

This leads to the soft-thresholding rule:

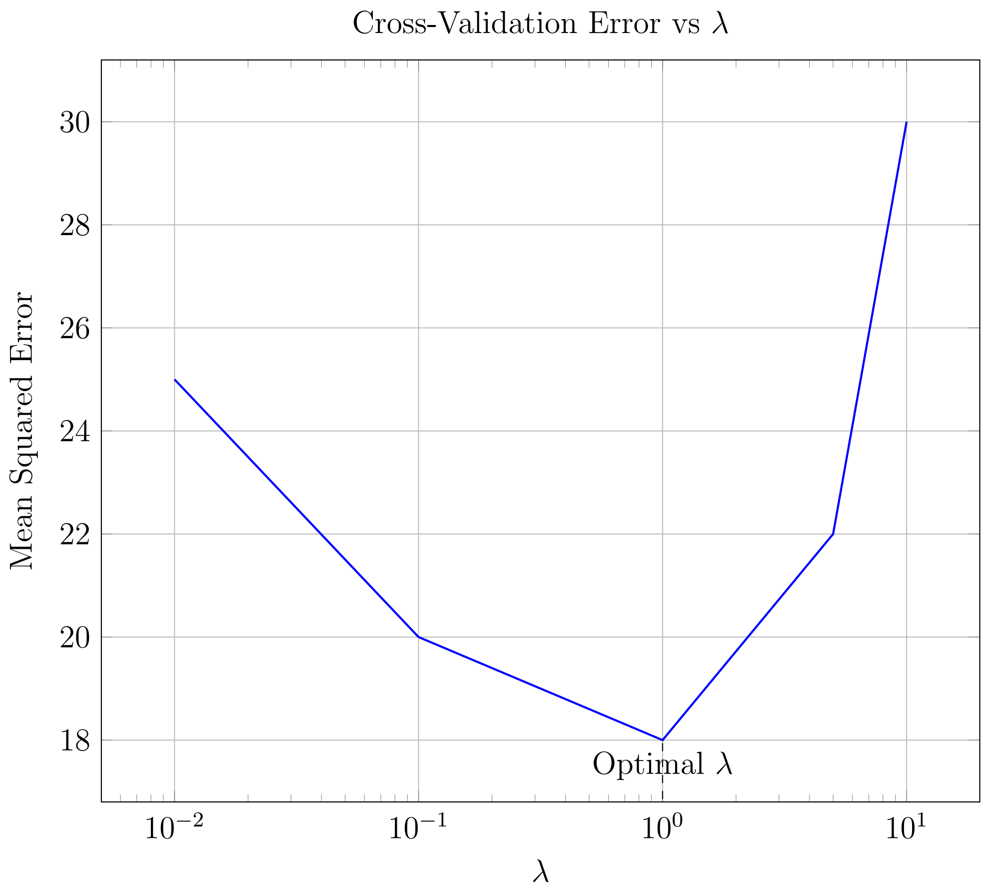
$$\beta_j = \begin{cases} \text{sign}(z)(|z| - \lambda) & \text{if } |z| > \lambda \\ 0 & \text{otherwise} \end{cases}$$

where z is the least squares solution.

4 Practical Implementation

4.1 Choosing λ

- Use cross-validation to select optimal λ
- Rule of thumb: Start with λ_{max} where all $\beta_j = 0$
- Decrease λ until model performance plateaus



4.2 Interpreting Results

- Non-zero coefficients: Important predictors
- Zero coefficients: Excluded features
- Magnitude: Relative feature importance