

## Ridge Regression (L2 Regularization)

### PROBLEM

Large coefficients  
lead to overfitting



### SOLUTION

Add penalty term  
for large coefficients

Cost Function:

$$RSS + \lambda \sum \beta_i^2$$

(L2 Penalty)



### Coefficient Shrinkage

Shrinks coefficients toward zero (but not exactly zero)



### Hyperparameter $\lambda$

Controls regularization strength



### Effect of $\lambda$

Larger  $\lambda \rightarrow$  more regularization, simpler model

### ✓ Key Benefit

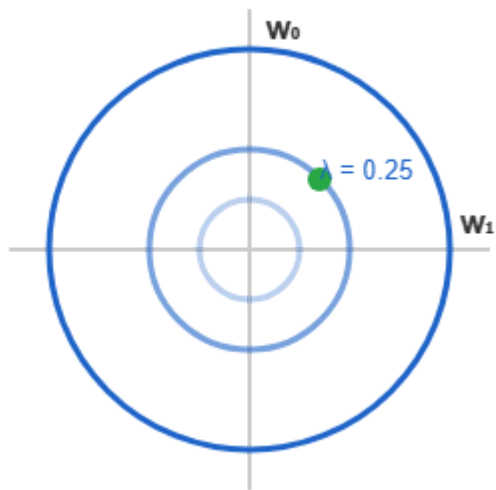
Helps with multicollinearity problem



## 2D Visual Understanding

L2 (Ridge) ●

L1 (Lasso) ●



### Circle shape

$$w_0^2 + w_1^2 = \text{constant}$$

Gradually shrinks coefficients



### Diamond shape

$$|w_0| + |w_1| = \text{constant}$$

Can make coefficients exactly zero



## Interactive: Effect of $\lambda$ on Coefficients

$\lambda$  (Lambda) = 0.0



💡 Move the slider to change  $\lambda$  value. As  $\lambda$  increases, coefficients shrink toward zero!

$\beta_1$  (original: 5.00)

$\beta_2$  (original: 3.00)

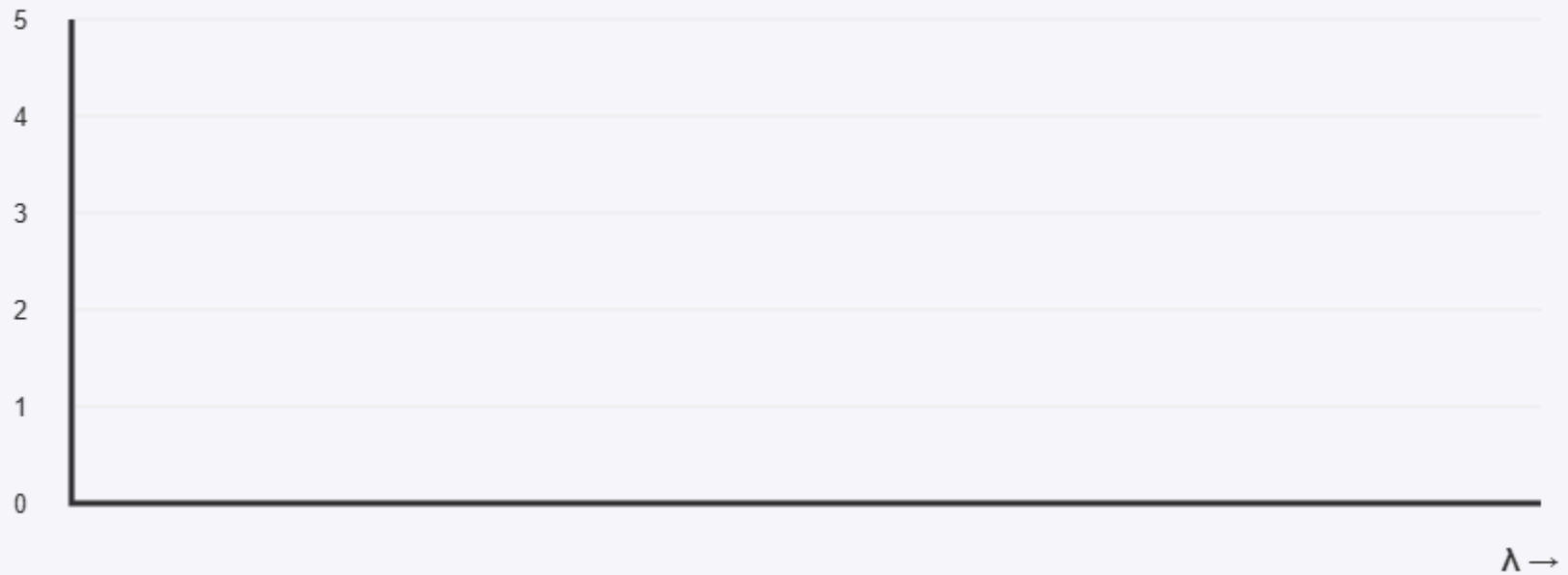
$\beta_3$  (original: 4.00)

5.00

3.00

4.00

Coefficient Value



## Geometric Intuition

Ridge regression finds the optimal solution by balancing two objectives:

### 1 Minimize Prediction Error (RSS)

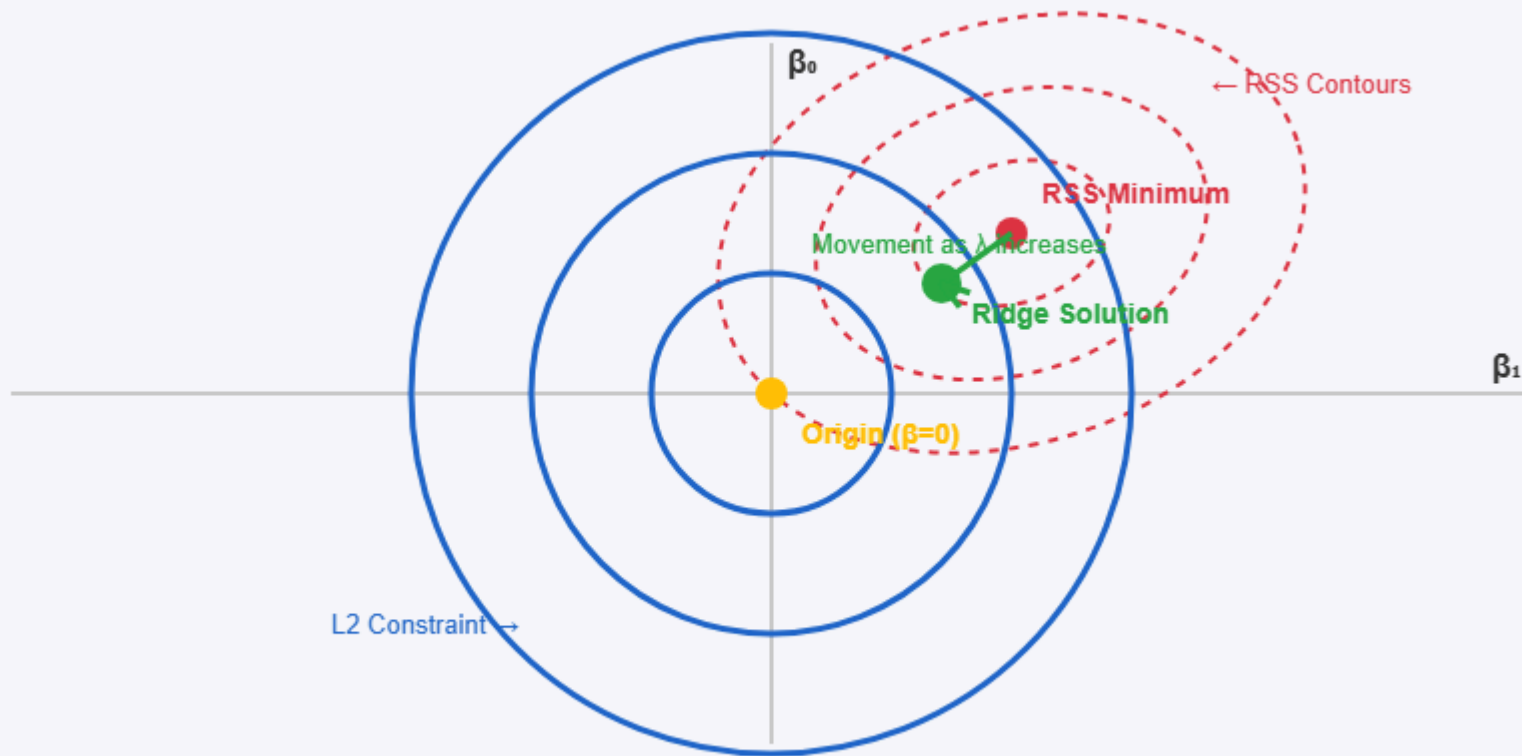
Want to fit the training data well


### 2 Minimize Coefficient Magnitude ( $\lambda \sum \beta^2$ )

Want to keep the model simple

$$\text{Loss} = \text{RSS} + \lambda \sum \beta_i^2$$

The solution moves from the RSS optimal point  
toward the origin ( $\beta=0$ ) as  $\lambda$  increases



 **Key Insight:** L2 penalty creates a circular constraint. The optimal solution is where the RSS contour (ellipse) touches the L2 constraint circle. As  $\lambda$  increases, the circle gets smaller, pulling coefficients toward zero.

 **Why Ridge Helps with Multicollinearity**

When features are highly correlated (multicollinear), ordinary linear regression can produce very large coefficients that cancel each other out.

 **Problem Example:**

If  $x_1 \approx x_2$ :

$$y = 100 \cdot x_1 - 99 \cdot x_2$$

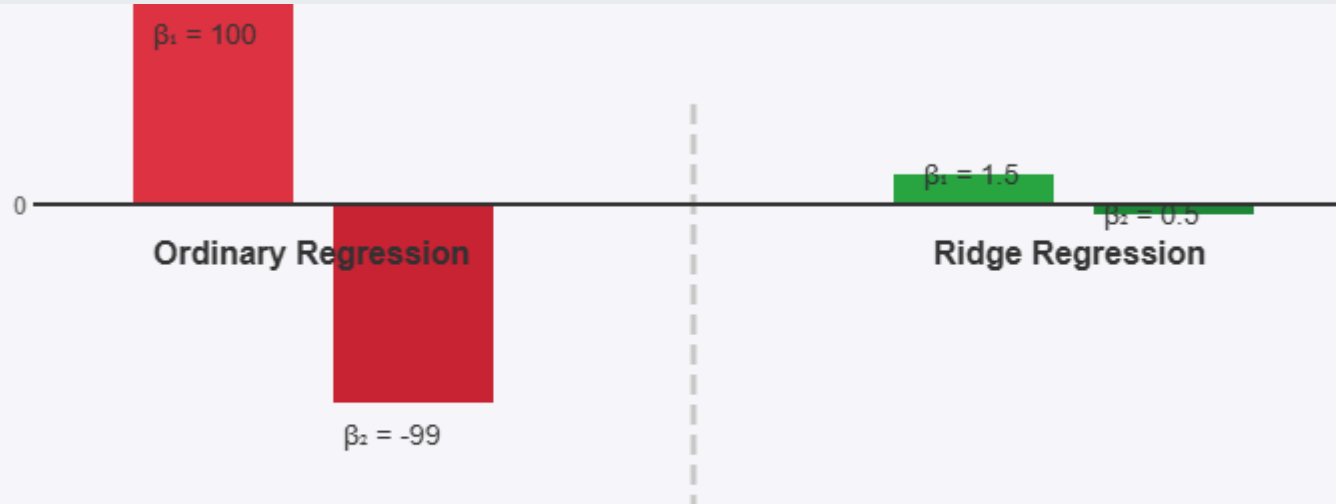
These large coefficients are unstable and sensitive to small data changes!

 **Ridge Solution:**

By penalizing large coefficients:

$$y = 1.5 \cdot x_1 + 0.5 \cdot x_2$$

The model becomes more stable and generalizes better!



**Coefficient Magnitude Comparison: With Multicollinearity**

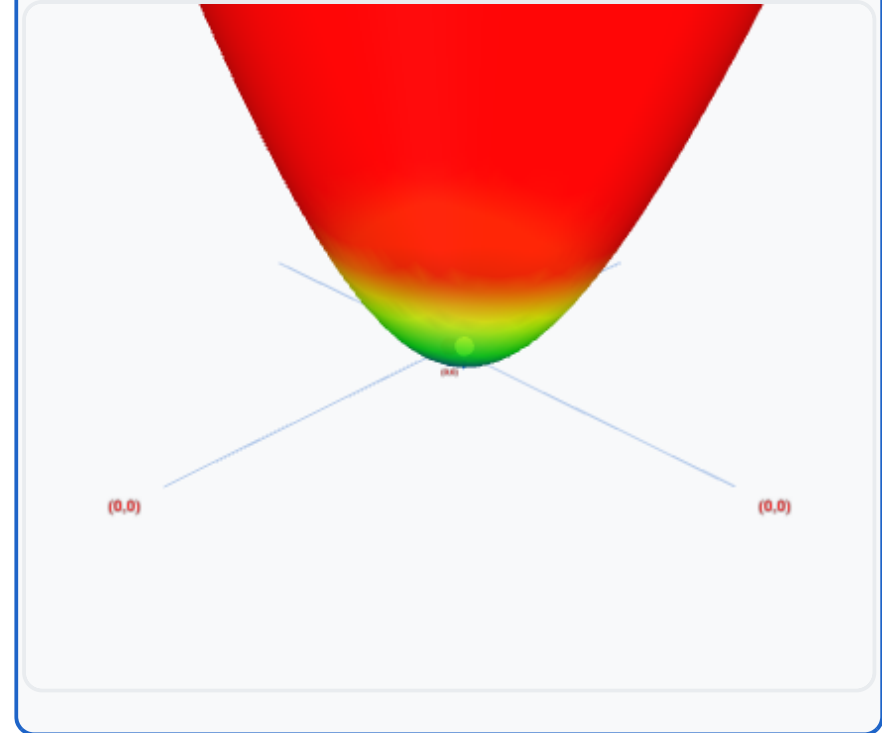
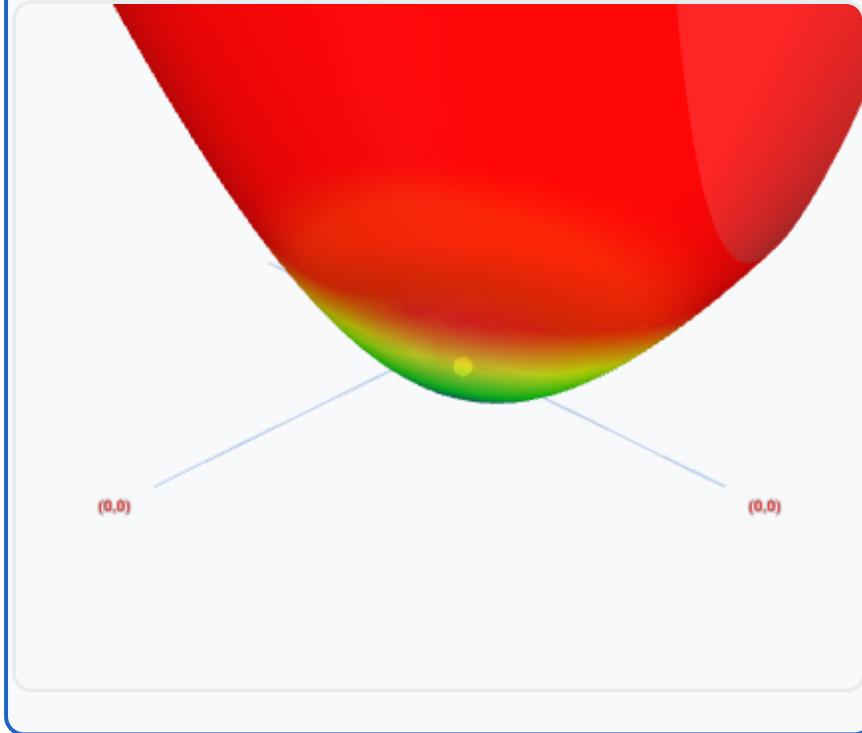


## 3D Loss Surface Visualization

Compare how the loss surface changes with and without L2 regularization.  
The white sphere marks the optimal coefficient values.

Ordinary Least Squares (OLS)

Ridge Regression (L2)




$\lambda$  (Lambda):



Rotation Speed:



 **Tip:** Increase  $\lambda$  to see how the optimal point (white sphere) moves closer to the origin (red sphere at center), representing smaller coefficient values and a simpler model.