Lasso Regression (L1 Regularization)

L1 Penalty:

$$RSS + \lambda \sum |\beta_i|$$

Sparse Solutions
Makes some coefficients exactly zero

Feature Selection
Automatically selects important features

Hyperparameter λ
Controls sparsity level

Interpretability
Creates simpler, more interpretable models

Ridge vs Lasso

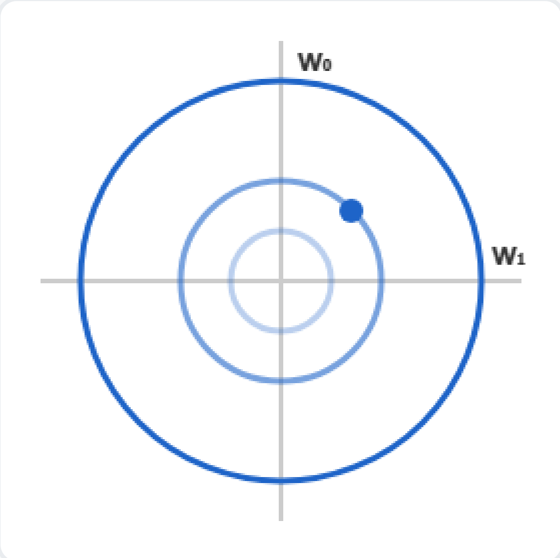
Ridge (L2) vs **Lasso (L1)**

Ridge shrinks coefficients, but Lasso can make them **exactly zero** → Feature selection!

💡 **Use Cross-Validation to choose optimal λ**

🎯 2D Visual Understanding

L2 (Ridge) 🔵



Circle shape

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

Gradually shrinks coefficients

L1 (Lasso) 🔴

Diamond shape

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

Can make coefficients exactly zero

🎚️ Interactive: Effect of λ on Lasso Coefficients

λ (Lambda) = 0.0

💡 Move the slider to change λ value. Notice how Lasso makes coefficients exactly zero!

β_1 (original: 5.00)	β_2 (original: 3.00)	β_3 (original: 4.00)
5.00	**3.00**	**4.00**

📐 Geometric Intuition: Why Lasso Creates Sparsity

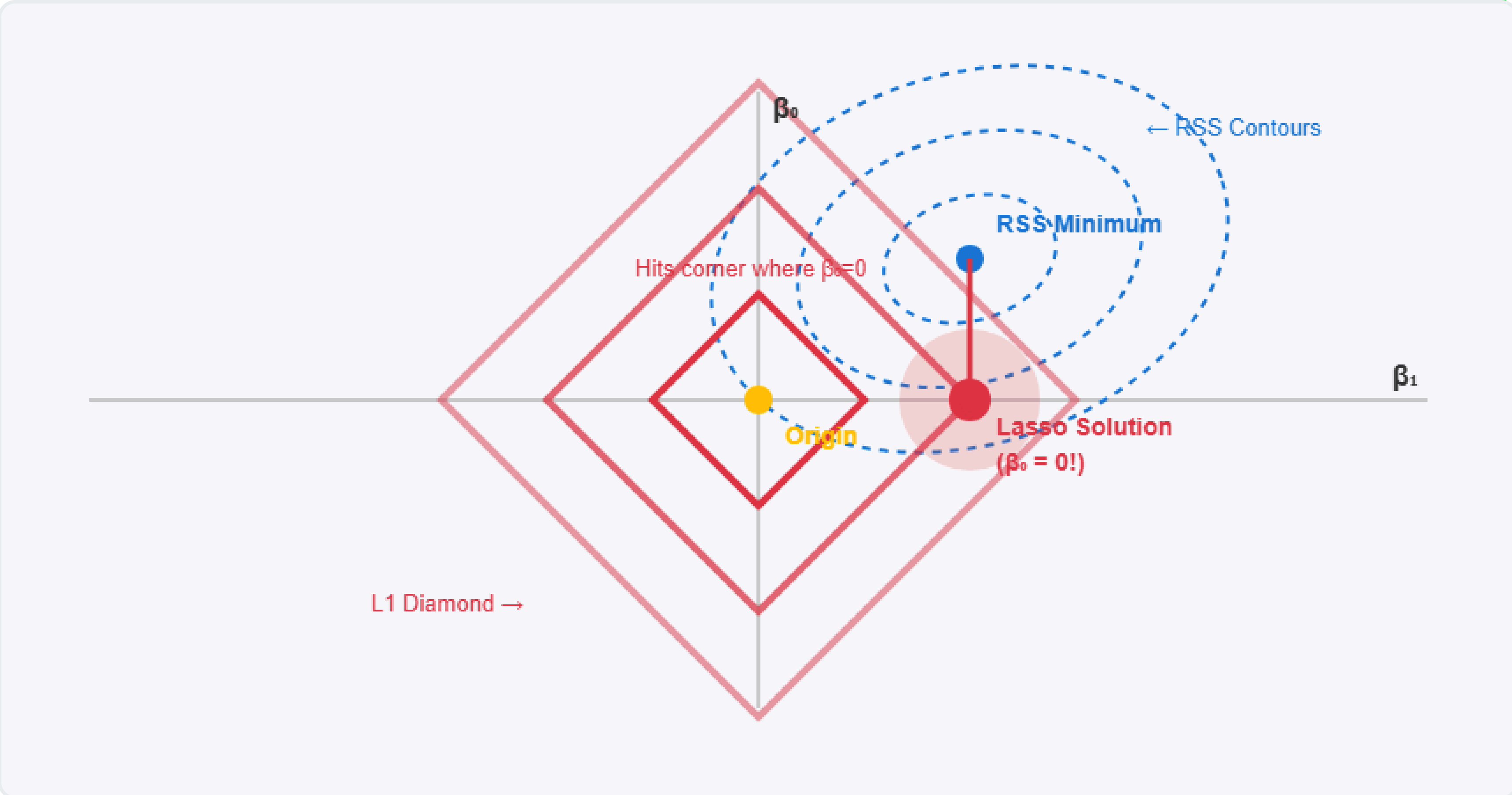
Lasso regression finds the optimal solution by balancing two objectives:

1️⃣ **Minimize Prediction Error (RSS)**
Want to fit the training data well

2️⃣ **Minimize Coefficient Magnitude ($\lambda\sum|\beta|$)**
Want to keep the model simple with sparse coefficients

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

The diamond shape of L1 constraint
has corners where coefficients become exactly zero



🎯 **Key Insight:** L1 penalty creates a diamond-shaped constraint. The optimal solution tends to hit the corners of the diamond, where one or more coefficients are exactly zero. This is why Lasso performs automatic feature selection!

🔍 Why Sparsity Matters: Feature Selection

When you have many features, not all of them are useful. Lasso automatically identifies and removes irrelevant features by setting their coefficients to zero.

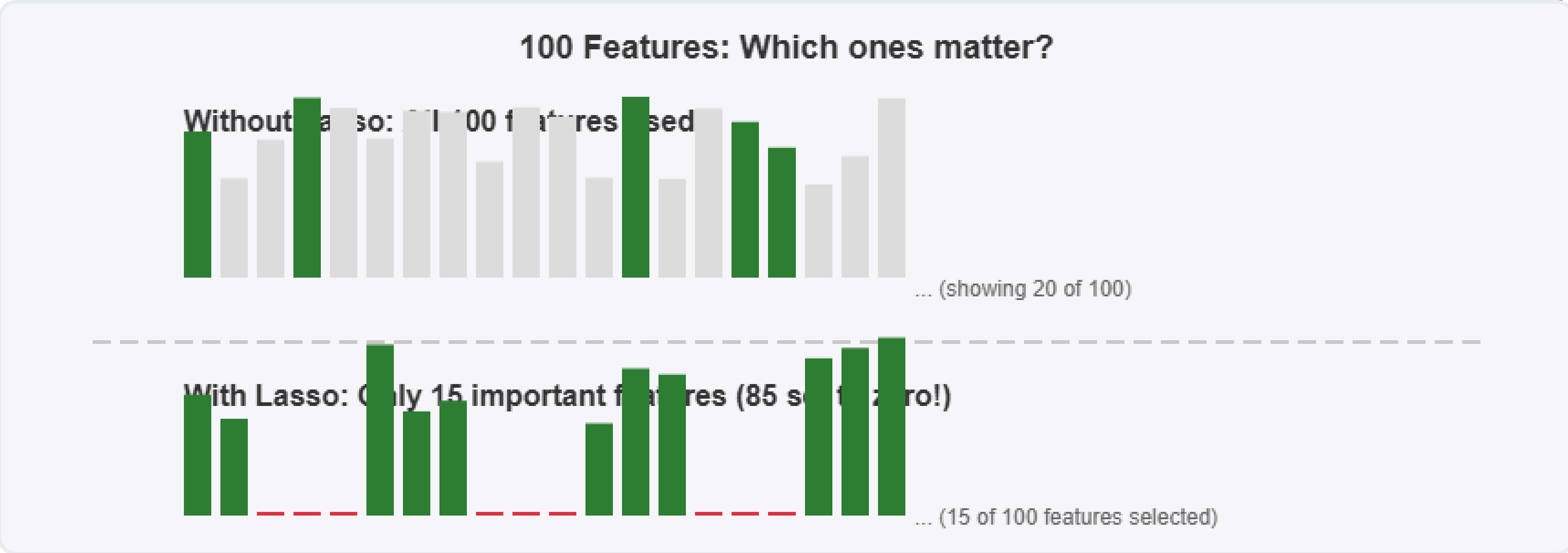
💼 **Real-world Scenario:**
Predicting house prices with 100 features:

```
Many features like "color of doorknob" are irrelevant
```

Lasso identifies that only 15 features truly matter and sets the rest to zero.

✅ Benefits of Sparsity:

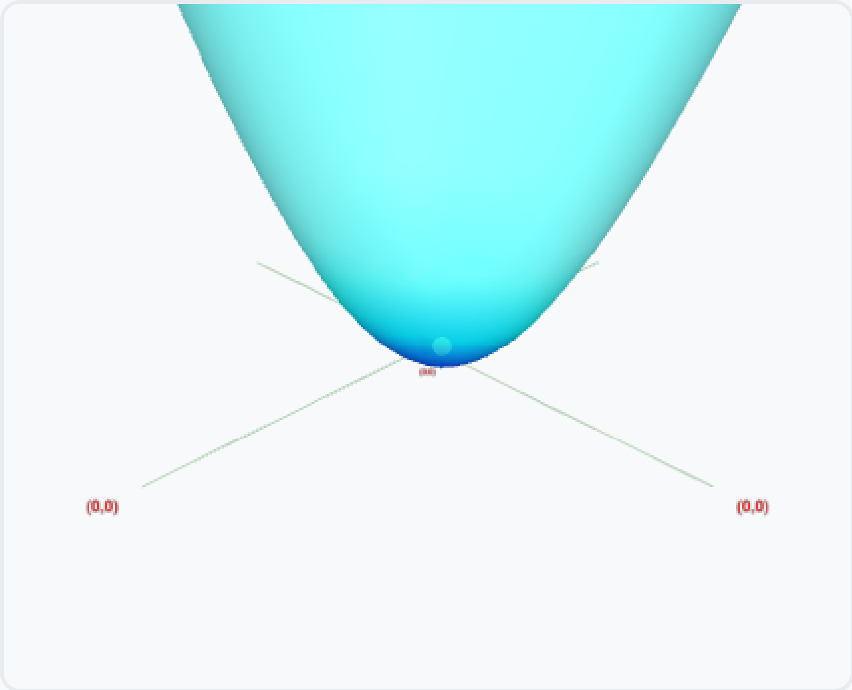
1. **Interpretability:** Simpler models are easier to understand
2. **Efficiency:** Fewer features mean faster predictions
3. **Generalization:** Removes noise, improving performance on new data
4. **Cost savings:** Don't need to collect/store irrelevant features

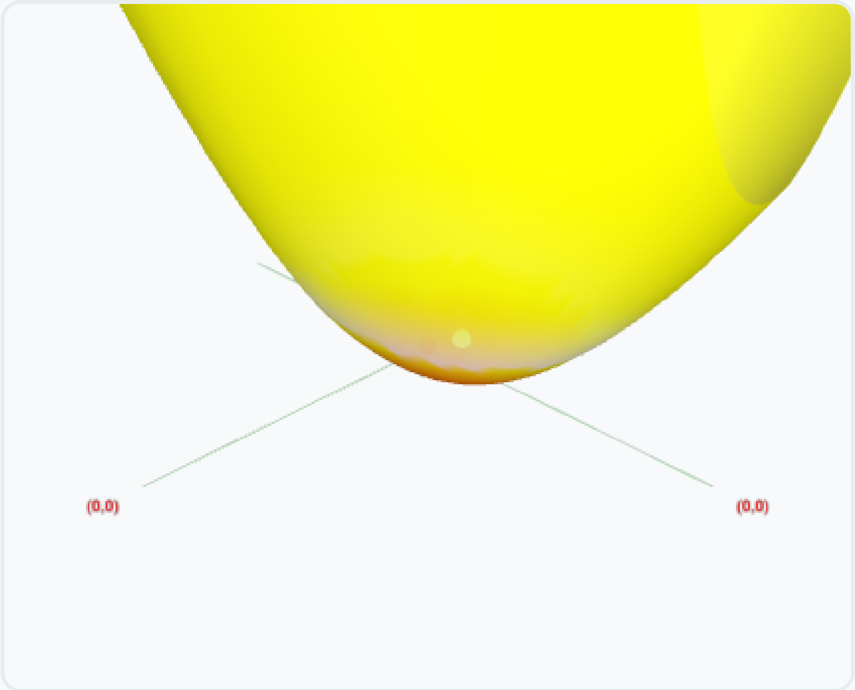


3D Loss Surface Visualization

Compare how the loss surface changes between Ridge (L2) and Lasso (L1) regularization.
Notice how Lasso's solution (white sphere) tends to align with axes, indicating zero coefficients.

Ridge Regression (L2)	Lasso Regression (L1)





λ (Lambda): 1.0

Rotation Speed: 1.0

💡 **Tip:** Increase λ to see how Lasso's optimal point (white sphere) moves toward the axes, while Ridge's point moves smoothly toward the origin. This visualizes why Lasso creates sparse solutions!