Comparison of Ridge and Lasso

CS109OA Introduction to Data Science

Pavlos Protopapas, Kevin Rader, and Chris Gumb

Photo: Jenny Zhu
Chicago

Ridge, LASSO - Computational complexity

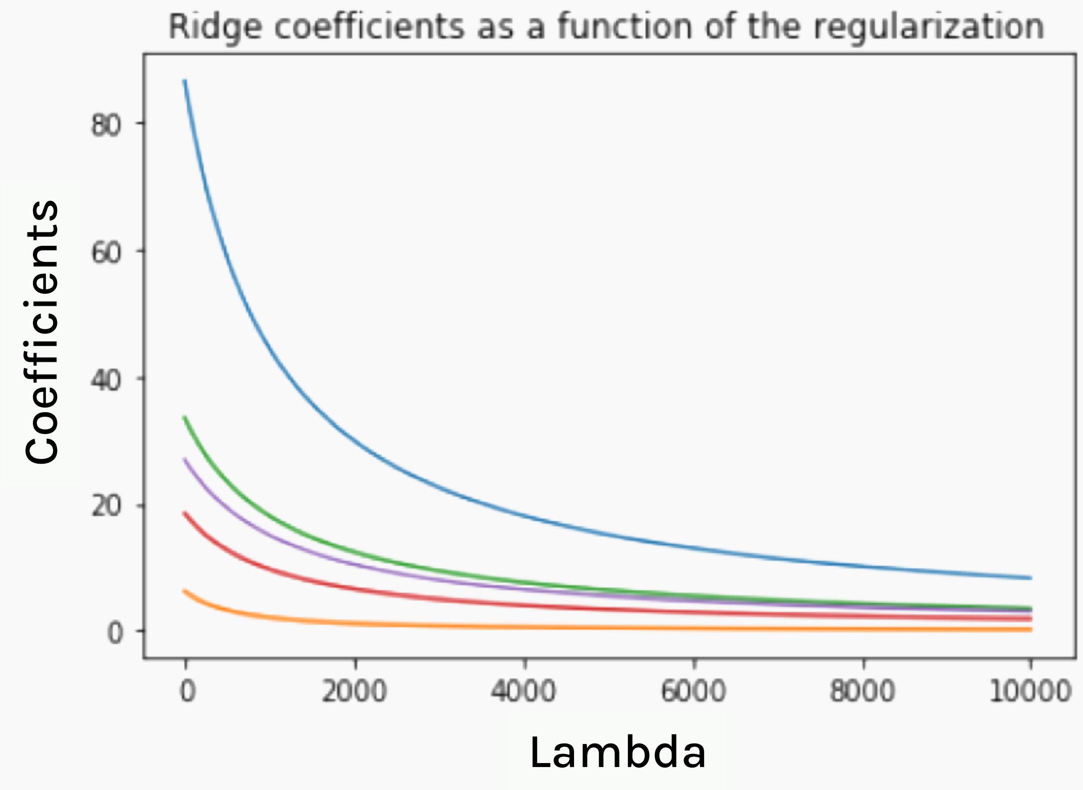
Solution to ridge regression:

$$\beta = (X^T X + \lambda I)^{-1} X^T Y$$

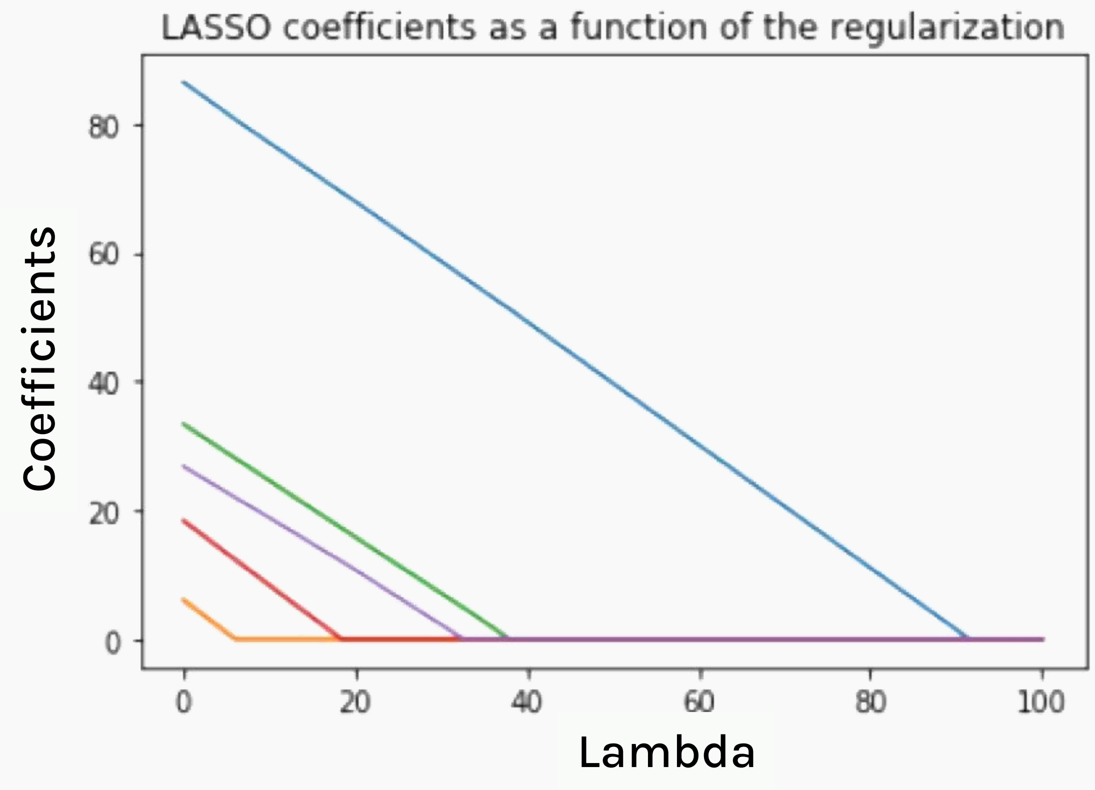
Solution to the LASSO regression:

LASSO has no conventional analytical solution, as the L1 norm has no derivative at zero. We can, however, use the concept of subdifferential or subgradient to find a manageable expression.

Ridge and LASSO visualized

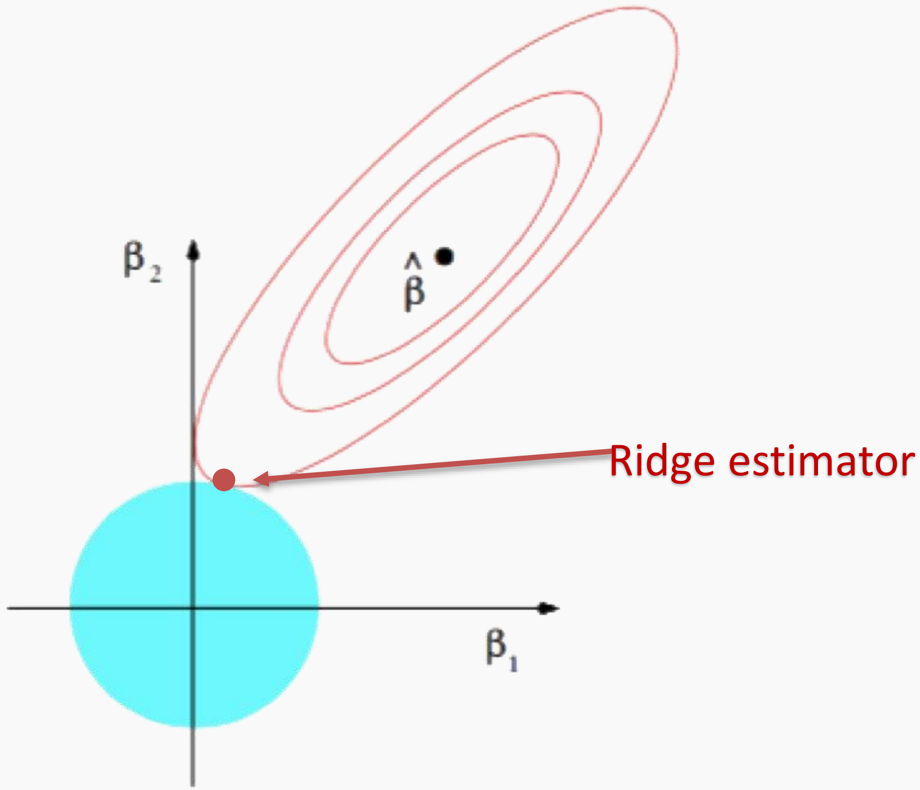


The values of the coefficients decrease as lambda increases, but they are not nullified.

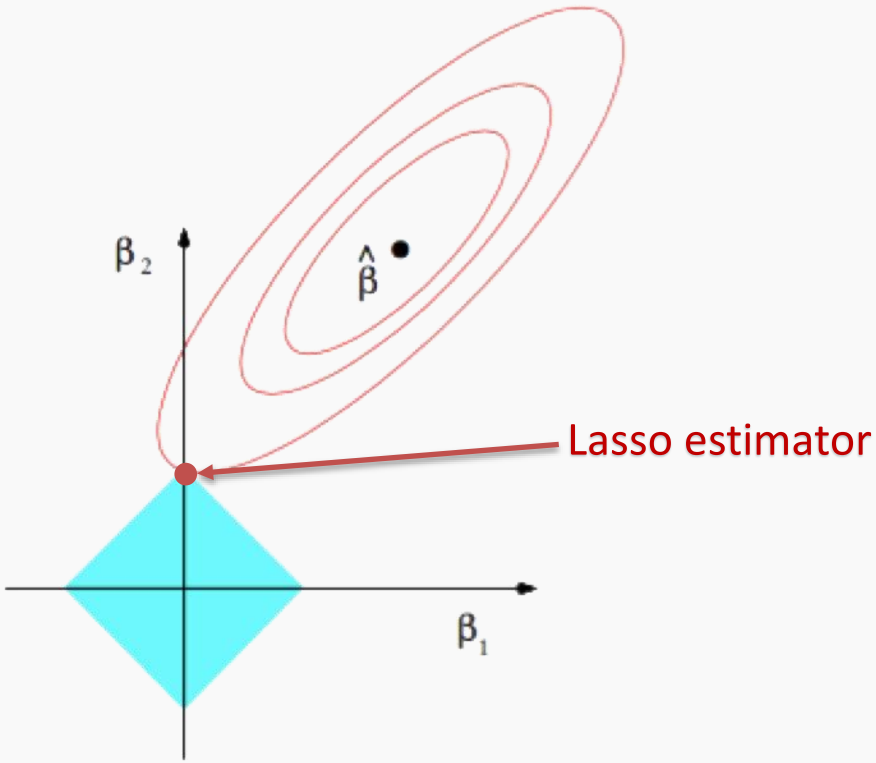


The values of the coefficients decrease as lambda increases and are nullified fast.

Ridge and LASSO visualized



The ridge estimator is where the constraint and the loss intersect.



The Lasso estimator tends to zero out parameters as the OLS loss can easily intersect with the constraint on one of the axis.

Comparing Lasso and Ridge Regression: Pros and Cons

Lasso Regression

Pros:

1. **Feature Selection**: Automatically performs feature selection by setting some coefficients to zero.
2. **Simplicity**: Results in simpler models that are easier to interpret.

Cons:

1. **Computational Complexity**: Can be computationally expensive when dealing with very large datasets.

Ridge Regression

Pros:

1. **Stability**: Stabilizes the coefficients, making it useful for ill-conditioned or multicollinear data.
2. **Computational Complexity**: Reduces model complexity, which can lead to more efficient computation and easier interpretation of results.

Cons:

1. **Feature Selection:** Does not reduce the number of features.

CS109A
Who Wants to be a Data Scientist?

Which of the following will give the correct β's in linear regression?

Options

A. $$\beta_1 = \sum_i (x_i\, y_i \;-\; x_i\, \bar{y} \;-\; \bar{x}\, y_i \;+ \bar{y})$$
$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

B. $$\beta = (X^T X)^{-1} X^T Y$$

C. $$\beta_1 = \sum_i (y_i \;- \beta_0) \,/\, \sum_i x_i$$
$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

D. $$\beta_1 = \sqrt{(x_{11} - x_{21})^2 + (x_{21} - x_{22})^2}$$
$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

Which of the following is **NOT** a purpose for model validation?

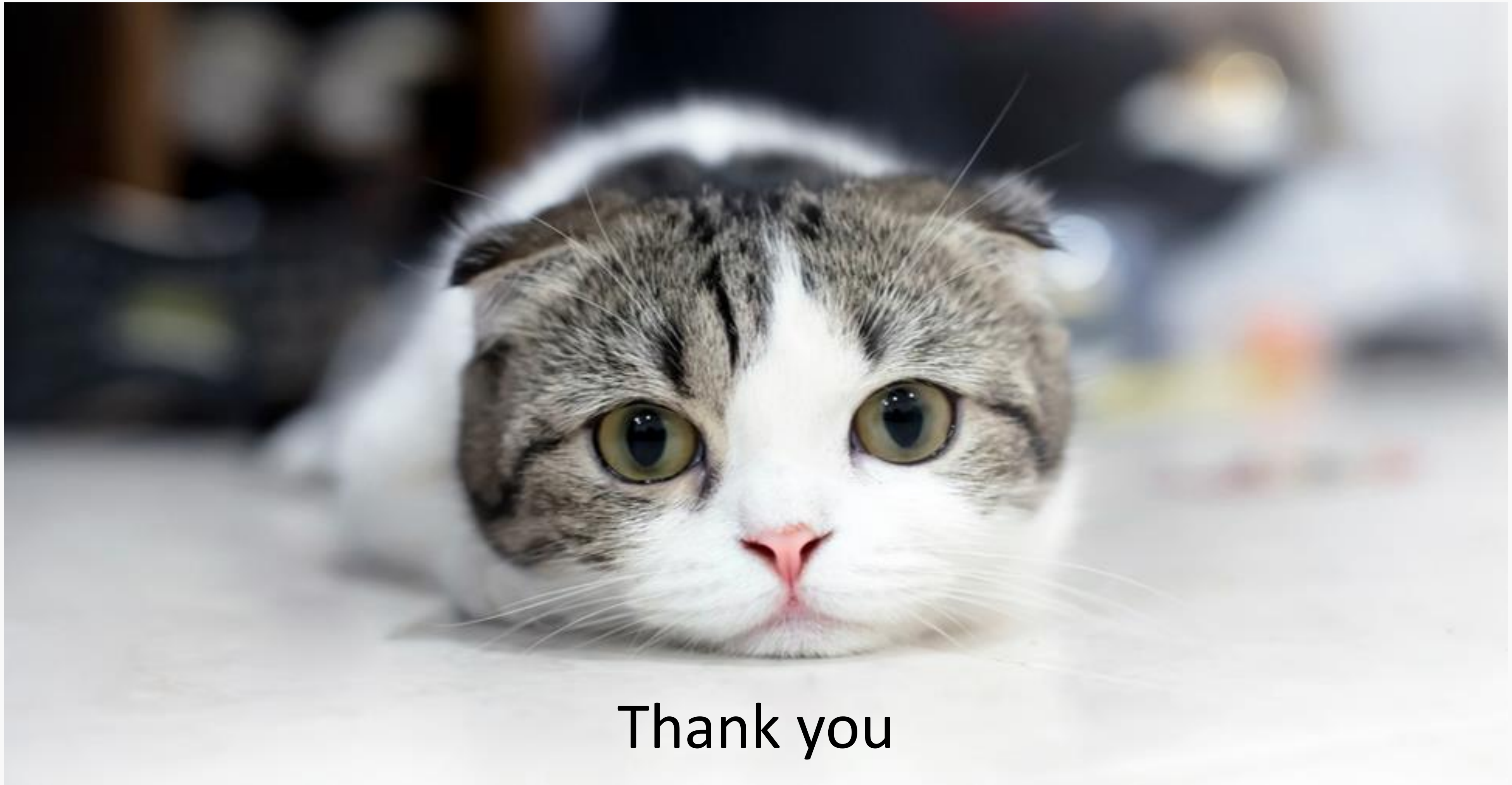
Options

A. To understand how well our model generalizes to data beyond the training set.

B. To determine the best combination of hyperparameters.

C. To measure, monitor, and hopefully reduce overfitting.

D. To test our model and report its final performance score.


Thank you