Comparison of Ridge and LASSO Regularization

Ridge and LASSO - Computational complexity

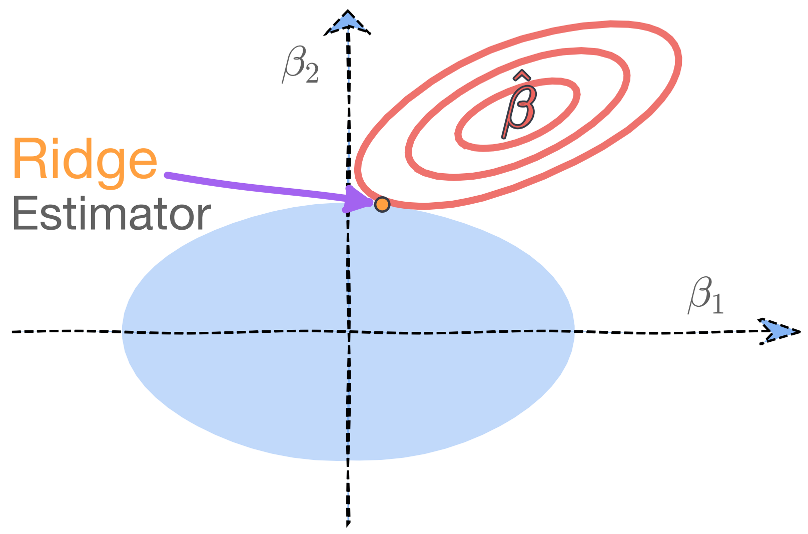
Solution to ridge regression:

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

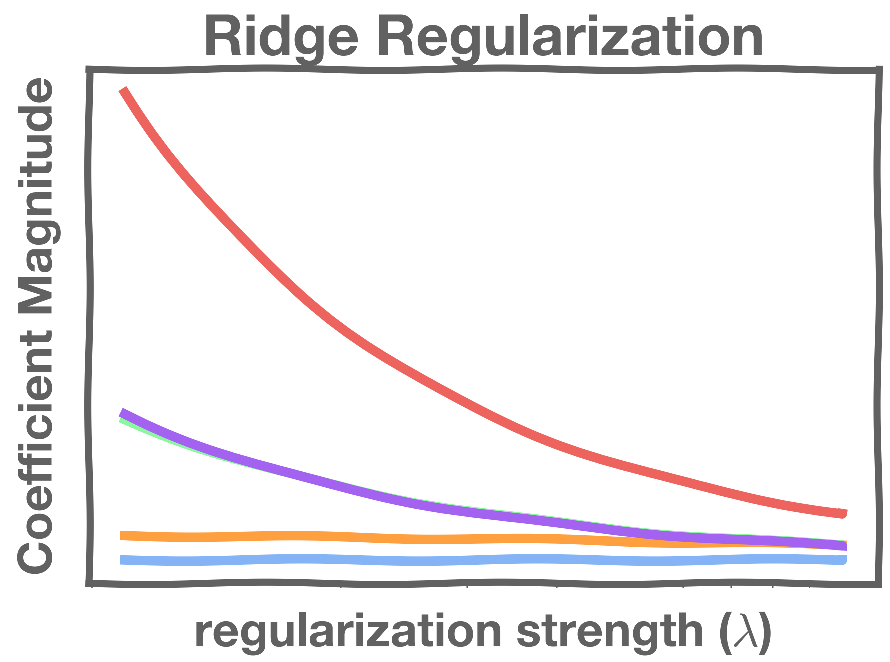
LASSO, on the other hand, *has no conventional analytical solution*, as the L1 norm has no derivative at zero. We can, however, use the concept of **subdifferential** or **subgradient** to find a manageable expression.

Ridge visualized

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

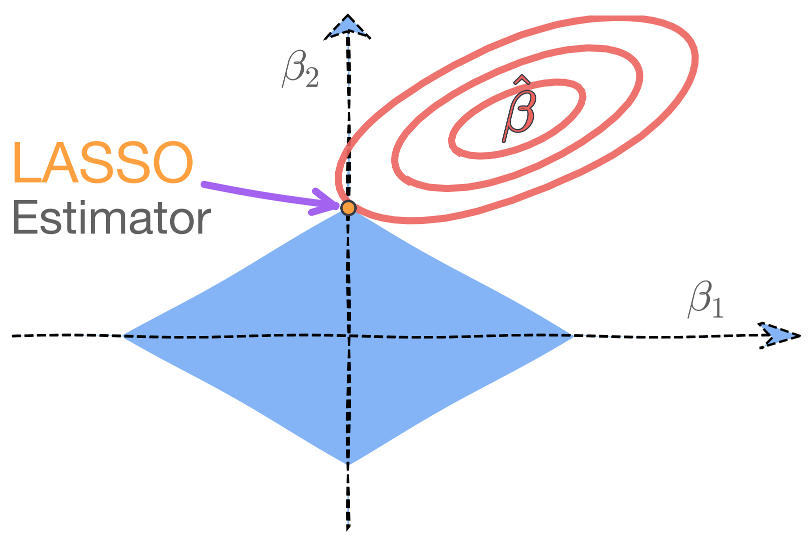


The values of the coefficients decrease as lambda increases, but they are not nullified.



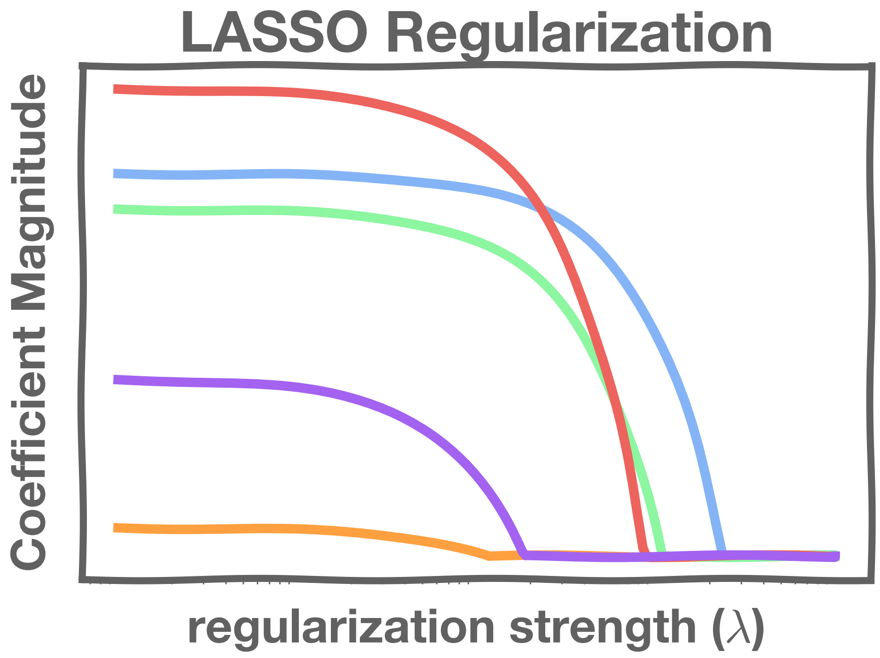
LASSO visualized

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



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

Calculator



VARIABLE SELECTION AS REGULARIZATION

What are the pros and cons of the two approaches?

Since **LASSO** regression tends to **produce zero estimates** for a number of model parameters - we say that LASSO solutions are **sparse** - we consider LASSO to be a **variable selection method**.

Ridge is **faster to compute**, but many people prefer using LASSO for variable selection, as well as for suppressing extreme parameter values and therefore being easier to **interpret**.

Discussion Board (External resource)

Haga clic en Aceptar para que su nombre de usuario y dirección de correo electrónico se envíen a una aplicación de terceros.

< Previous Next >

Calculator