

# LASSO & Ridge Regression - Everything you wanted to know

The origin of the term “Regularization”

<https://stats.stackexchange.com/questions/250722/the-origin-of-the-term-regularization?noredirect=1&lq=1>

Why Does Shrinkage work?

<https://stats.stackexchange.com/questions/179864/why-does-shrinkage-work>

What problem do Shrinkage methods solve?

<https://stats.stackexchange.com/questions/20295/what-problem-do-shrinkage-methods-solve?noredirect=1&lq=1>

Regression Shrinkage Methods

<https://online.stat.psu.edu/stat508/lesson/5>

Regression Shrinkage and Selection via the LASSO

[https://www.ccs.neu.edu/home/eelhami/courses/EE290A/LASSO Tibshirani.pdf](https://www.ccs.neu.edu/home/eelhami/courses/EE290A/LASSO_Tibshirani.pdf)

Why does the LASSO provide Variable Selection?

<https://stats.stackexchange.com/questions/74542/why-does-the-lasso-provide-variable-selection>

Understanding LASSO Regression’s sparsity geometrically

<https://stats.stackexchange.com/questions/414491/understanding-lasso-regressions-sparsity-geometrically?rq=1>

Why is Ridge Regression called “Ridge”, why is it needed and what happens when  $\lambda$  goes to infinity?

<https://stats.stackexchange.com/questions/151304/why-is-ridge-regression-called-ridge-why-is-it-needed-and-what-happens-when?noredirect=1&lq=1>

What are disadvantages of using the LASSO for variable selection for Regression?

<https://stats.stackexchange.com/questions/7935/what-are-disadvantages-of-using-the-lasso-for-variable-selection-for-regression>

Ridge vs LASSO (Statquest series)

[https://www.youtube.com/watch?v=Xm2C\\_gTAI8c&ab\\_channel=StatQuestwithJoshStarmer](https://www.youtube.com/watch?v=Xm2C_gTAI8c&ab_channel=StatQuestwithJoshStarmer)

[https://www.youtube.com/watch?v=Q81RR3yKn30&t=770s&ab\\_channel=StatQuestwithJoshStarmer](https://www.youtube.com/watch?v=Q81RR3yKn30&t=770s&ab_channel=StatQuestwithJoshStarmer)

[https://www.youtube.com/watch?v=NGf0voTmlcs&ab\\_channel=StatQuestwithJoshStarmer](https://www.youtube.com/watch?v=NGf0voTmlcs&ab_channel=StatQuestwithJoshStarmer)

Why LASSO or ElasticNet perform better than Ridge when features are correlated

<https://stats.stackexchange.com/questions/264016/why-lasso-or-elasticnet-perform-better-than-ridge-when-the-features-are-correlated?noredirect=1&lq=1>

Detailed explanation of LASSO and RIDGE in “The Elements of Statistical Learning”

<https://hastie.su.domains/ElemStatLearn/download.html>

Why use Lasso estimates over OLS estimates on the Lasso-identified subset of variables?

<https://stats.stackexchange.com/questions/82466/why-use-lasso-estimates-over-ols-estimates-on-the-lasso-identified-subset-of-variables>

Is Bayesian ridge regression another name of Bayesian Linear Regression?

<https://stats.stackexchange.com/questions/328342/is-bayesian-ridge-regression-another-name-of-bayesian-linear-regression?noredirect=1&lq=1>

When should I use LASSO vs Ridge?

<https://stats.stackexchange.com/questions/866/when-should-i-use-lasso-vs-ridge?noredirect=1&lq=1>

If only prediction is of interest, why use LASSO over Ridge?

<https://stats.stackexchange.com/questions/331782/if-only-prediction-is-of-interest-why-use-lasso-over-ridge?noredirect=1&lq=1>

Is regression with L1 regularization the same as Lasso, and with L2 regularization the same as ridge regression? And how to write "Lasso"?

<https://stats.stackexchange.com/questions/200416/is-regression-with-l1-regularization-the-same-as-lasso-and-with-l2-regularizati?noredirect=1&lq=1>

Why can't ridge regression provide better interpretability than LASSO ?

<https://stats.stackexchange.com/questions/179611/why-cant-ridge-regression-provide-better-interpretability-than-lasso?rq=1>

Regression in  $p > n$  setting: how to choose regularization method (Lasso, PLS, PCR, ridge)?

<https://stats.stackexchange.com/questions/108614/regression-in-pn-setting-how-to-choose-regularization-method-lasso-pls-pc?rq=1>

What is the advantage of reducing dimensionality of predictors for the purposes of regression?

<https://stats.stackexchange.com/questions/36249/what-is-the-advantage-of-reducing-dimensionality-of-predictors-for-the-purposes?noredirect=1&lq=1>

Relationship between Ridge Regression and PCA Regression

<https://stats.stackexchange.com/questions/81395/relationship-between-ridge-regression-and-pca-regression?noredirect=1&lq=1>

Is Tikhonov Regularization the same as Ridge Regression?

<https://stats.stackexchange.com/questions/234280/is-tikhonov-regularization-the-same-as-ridge-regression?rq=1>

Why do we only see L1 and L2 regularization but not other norms?

<https://stats.stackexchange.com/questions/269298/why-do-we-only-see-l-1-and-l-2-regularization-but-not-other-norms?noredirect=1&lq=1>

What are the assumptions of Ridge Regression and how to test them?

<https://stats.stackexchange.com/questions/169664/what-are-the-assumptions-of-ridge-regression-and-how-to-test-them?rq=1>

Under exactly What conditions is ridge regression able to provide an improvement over ordinary least squares regression?

<https://stats.stackexchange.com/questions/122936/under-exactly-what-conditions-is-ridge-regression-able-to-provide-an-improvement?rq=1>

Interpretation of Ridge Regularization in Regression.

<https://stats.stackexchange.com/questions/130017/interpretation-of-ridge-regularization-in-regression?rq=1>

How to find regression coefficients Beta in Ridge Regression?

<https://stats.stackexchange.com/questions/190993/how-to-find-regression-coefficients-beta-in-ridge-regression?noredirect=1&lq=1>

The limit of 'Unit-variance' Ridge Regression estimator when  $\lambda \rightarrow \infty$

<https://stats.stackexchange.com/questions/340099/the-limit-of-unit-variance-ridge-regression-estimator-when-lambda-to-infty?rq=1>

When would least squares be a bad idea?

<https://stats.stackexchange.com/questions/82864/when-would-least-squares-be-a-bad-idea?rq=1>

Why does Ridge estimate become better than OLS by adding a constant to the diagonal?

<https://stats.stackexchange.com/questions/118712/why-does-ridge-estimate-become-better-than-ols-by-adding-a-constant-to-the-diago?rq=1>

Is Ridge Regression useless in high dimensions ( $n < p$ ) ? How can OLS fail to overfit?

<https://stats.stackexchange.com/questions/328630/is-ridge-regression-useless-in-high-dimensions-n-lt-p-how-can-ols-fail-to?noredirect=1&lq=1>

Is Fig 3.6 in Elements of Statistical Learning Correct?

<https://stats.stackexchange.com/questions/496539/is-fig-3-6-in-elements-of-statistical-learning-correct?noredirect=1&lq=1>

Algorithms for automatic model selection

<https://stats.stackexchange.com/questions/20836/algorithms-for-automatic-model-selection?noredirect=1&lq=1>

Unified view on shrinkage: what is the relations (if any) between Stein's paradox, ridge regression, and random effects in mixed models?

<https://stats.stackexchange.com/questions/122062/unified-view-on-shrinkage-what-is-the-relation-if-any-between-steins-paradox?noredirect=1&lq=1>

Quora question: what is the difference between L1 and L2 regularization? How does it solve the problem of overfitting? Which regularizer to use and when?

<https://www.quora.com/What-is-the-difference-between-L1-and-L2-regularization-How-does-it-solve-the-problem-of-overfitting-Which-regularizer-to-use-and-when>

Aryma Labs