

Regularization Techniques

Three sources of error for your model are: bias, variance, and, irreducible error.

Regularization is a way to achieve building simple models with relatively low error. It helps you avoid overfitting by penalizing high-valued coefficients. It reduces parameters and shrinks the model.

Regularization adds an adjustable regularization strength parameter directly into the cost function.

Regularization performs feature selection by shrinking the contribution of features, which can prevent overfitting.

In Ridge Regression, the complexity penalty λ is applied proportionally to squared coefficient values.

- The penalty term has the effect of “shrinking” coefficients toward 0.
- This imposes bias on the model, but also reduces variance.
- We can select the best regularization strength λ via cross-validation.
- It's a best practice to scale features (i.e. using `StandardScaler`) so penalties aren't impacted by variable scale.

In LASSO regression: the complexity penalty λ (λ) is proportional to the absolute value of coefficients. LASSO stands for : Least Absolute Shrinkage and Selection Operator.

- Similar effect to Ridge in terms of complexity tradeoff: increasing λ raises bias but lowers variance.
- LASSO is more likely than Ridge to perform feature selection, in that for a fixed λ , LASSO is more likely to result in coefficients being set to zero.

Elastic Net combines penalties from both Ridge and LASSO regression. It requires tuning of an additional parameter that determines emphasis of L1 vs. L2 regularization penalties.

LASSO's feature selection property yields an interpretability advantage, but may underperform if the target truly depends on many of the features.

Elastic Net, an alternative hybrid approach, introduces a new parameter α (alpha) that determines a weighted average of L1 and L2 penalties.

Regularization techniques have an analytical, a geometric, and a probabilistic interpretation.