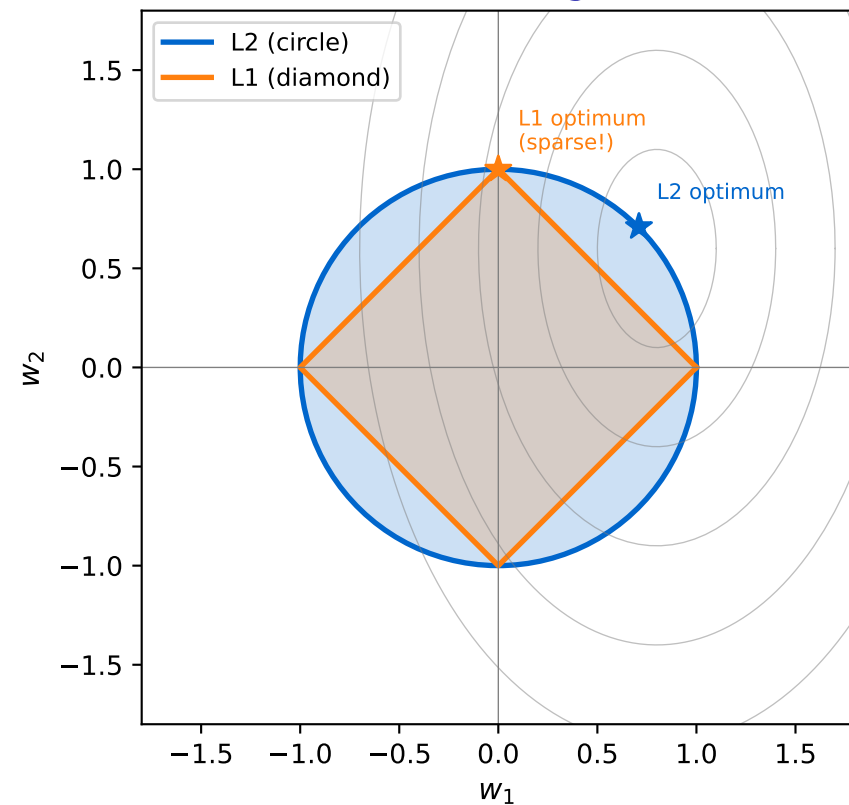
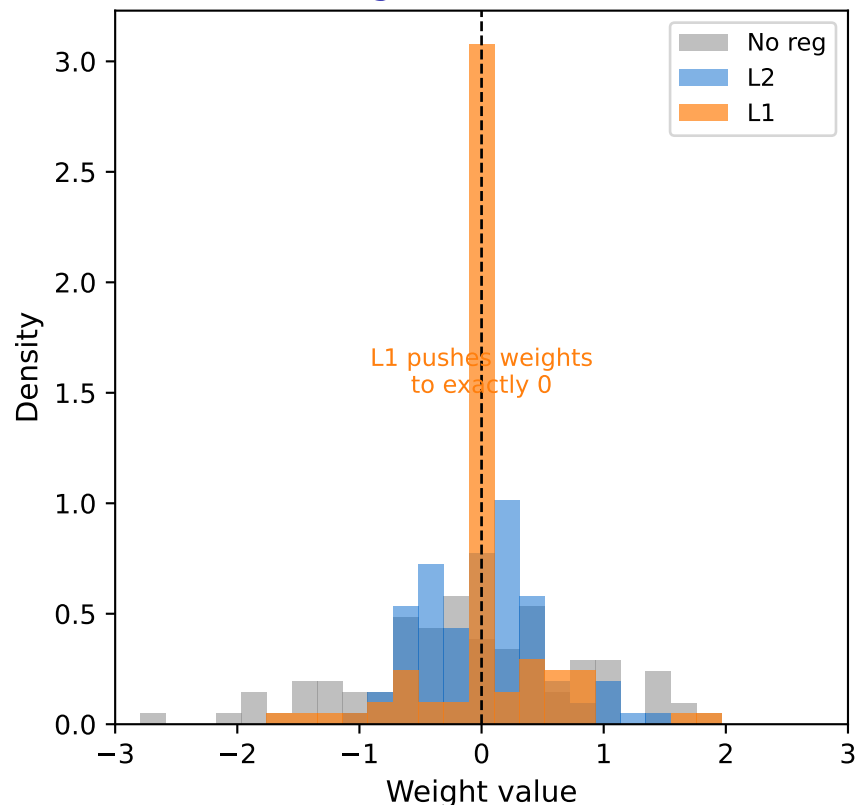


L1 vs L2: Different Regularization Philosophies

Constraint Regions



Weight Distributions



L1 vs L2 Regularization

	L1 (Lasso)	L2 (Ridge)
Penalty	$\lambda \sum w_i $	$\lambda \sum w_i^2$
Effect	Sparse weights	Small weights
Feature	Feature selection	All features used
Geometry	Diamond	Circle
Use case	Many irrelevant features	All features matter

Elastic Net combines both: $\lambda_1 \sum |w_i| + \lambda_2 \sum w_i^2$