
Use LASSO when:

1. You expect only a few features to matter (sparse model)

If you believe the true relationship uses a **small subset** of variables, lasso is ideal.

✓ Good when many variables are irrelevant

✓ Automatically selects features by setting coefficients **to zero**

Example cases:

- Genetic data (thousands of genes, few are truly important)
- High-dimensional datasets ($p \gg n$)
- Situations where interpretability matters — "Which variables matter?"

---

2. You want a model that is easy to interpret

Lasso tells you which features the model uses.
All coefficients except a few become exactly zero.

✓ Useful for scientific modeling

✓ Useful for finance, medicine, social science

✓ Good for presentations where audiences need a simple formula

---

3. Your features are not strongly correlated

Lasso struggles when features are highly correlated:
it randomly picks one and sets the others to zero.
If correlations are high → Ridge or Elastic Net is better.

Use RIDGE when:

1. Features are highly correlated

Ridge shrinks correlated predictors together in a *smooth way*.

Example:

- Height and arm span
- x_1 and x_2 are nearly duplicates
- Many related features in multicollinearity situations

Ridge handles this beautifully; Lasso does not.

---

2. You expect many small/medium effects, not sparsity

If you think all variables matter *a little*, ridge is better.

✓ Keeps all variables

✓ Smooth shrinkage

✓ No hard selection

Example:

- Tikhonov regularization in ill-posed problems
- Continuous, smooth phenomena (physics, engineering)
- Data where every feature carries some information

---

3. Prediction accuracy is the priority (not interpretation)

Ridge tends to perform slightly better than lasso when:

- features are correlated
- all features have some signal
- noise is moderate

Elastic Net (the middle ground)

Sometimes you want:

✓ some variable selection

✓ but also good performance with correlated features

Elastic Net combines L1 + L2 penalties:

$$\lambda_1 |\beta|_1 + \lambda_2 |\beta|_2^2$$

It:

- selects features (like lasso)
- handles correlation (like ridge)

Often the **best practical choice**.

Quick rule-of-thumb summary

Situation	Best Choice	Reason
Many features, only a few important	**Lasso**	Selects them automatically
Highly correlated features	**Ridge**	Shares weight smoothly
$p \gg n$ (more features than samples)	**Lasso** or Elastic Net	Dimensionality reduction
Want interpretability	**Lasso**	Zero coefficients
All features have small effects	**Ridge**	Smooth shrinkage
Unsure	**Elastic Net**	Balanced behavior

One-sentence intuition

- **Lasso:** “Pick the important features and kill the rest.”
- **Ridge:** “Use all features but shrink them to avoid overfitting.”
