

Subset Selection

Forward Subset Selection

- Start with most explaining variable
- Add variables until it doesn't improve
- Prevents overfitting

Advanced Selections

- Ensemble selection
- Use multiple algorithms
- Use a starting pool

Backward Subset Selection

- Start with all variables
- Remove until no improvement
- Expected to perform better than forward (worse when noise is too much)

Forward & Backward

- Add or remove in any of steps

Mix Stepwise Selection

- Let the algorithm try every method till it finds a local optimal variable base

Subset Selection Penalty Metrics

• Cp
• BIC
find min

- AIC → uses mLE
- Adjusted R² → larger the better

Shrinkage

- Fit a model with p predictors and shrink their coefficients to zero.
- Shrinking reduces variance (prevents overfitting)

Ridge (L2)

$$\text{Penalty} = \text{RSS} + \lambda \sum_{j=1}^p \beta_j^2$$

small when $\beta_j \neq 0$ shrinkage penalty

Disadvantage → none of predictor coefficients go zero

tuning param for regularization

Lasso (L1)

$$\text{Penalty} = \text{RSS} + \lambda \sum_{j=1}^p |\beta_j|$$

Some coefficients go completely zero (like subset selection) if λ is sufficiently large
Yields sparse models

ElasticNet (L1+L2)

Lasso + Ridge
Can be given weights

No free lunch!