

Notes on Elem Stat Learn Ch. 3

September 6, 2022

1 Subset Selection

Forward/Backward selection approaches, based on measure of model fit

2 Ridge

$$\hat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (1)$$

Key points on Lasso

- Standardize all inputs, always
- “Ridge regression protects against the potentially high variance of gradients estimated in the short directions.”
- Ridge downweights everything

3 Lasso

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (2)$$

- The lasso does a kind of continuous subset selection
- Effectively the lasso pushes harder toward 0, but for fewer variables

4 Elastic Net

$$\lambda \sum_{j=1}^p (\alpha \beta_j^2 + (1 - \alpha) |B_j|) \quad (3)$$

- $\alpha = 1$ gives us ridge, $\alpha = 0$ gives us lasso, λ is the penalty.