Regularized regression II

Elastic Net and Group Lasso

Elastic net

A compromise between ridge and lasso

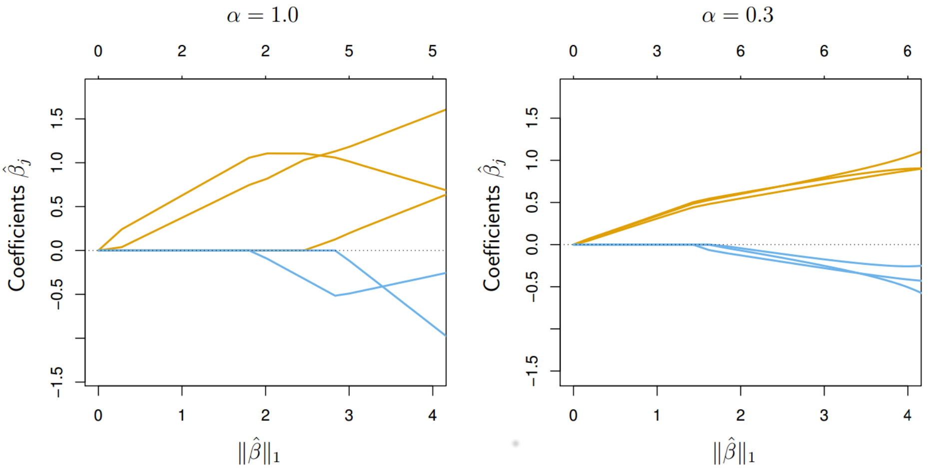
$$\underset{\beta_0,\beta}{\text{minimize}} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - x_i'\beta)^2 + \lambda \Big[\, (1-\alpha)\|\beta\|_2^2 + \alpha\|\beta\|_1 \Big] \right\}$$

Elastic net

- Introduces a mixing parameter $\alpha \in \big[0, 1\big]$
  - $\alpha = 0$: Ridge regression
  - $\alpha = 1$: Lasso
- α is an additional tuning parameter

Elastic net

Figure: Coefficient paths with elastic net



Hastie et al. (2015)

Extensions

Categorical outcomes

$$\underset{\beta_0,\beta}{\text{minimize}}\left\{-\frac{1}{n}\mathcal{L}(\beta_0,\beta;\mathbf{y},\mathbf{X})+\lambda\|\beta\|_1\right\}$$

- Shrinkage can be applied with binary and multinomial outcomes...
- ...by introducing a penalty in the likelihood function

A note on standardization

- Contribution to penalty term dependent on scale
- Ridge and Lasso typically applied with standardized features

Extensions

Penalty factors

$$\lambda \sum_{j=1}^{p} \upsilon_j P_\alpha(\beta_j) = \lambda \sum_{j=1}^{p} \upsilon_j \left[(1-\alpha)\frac{1}{2}\beta_j^2 + \alpha|\beta_j| \right]$$

- Allows to control the regularization process given prior/ substantive knowledge
- Specify separate penalty factors for individual coefficients
- Set υ_j equal to zero to not penalize at all

Group Lasso

Regularization with feature groups

- Standard lasso considers predictors independently
- Not desirable when features have a natural group structure
  - Groups of dummy variables
  - Main effects and interactions

$\rightarrow$ Group lasso in- or excludes groups of variables together

Group Lasso

Regression with covariate groups

- z_{ij} represents covariates in group j
- θ_j a group of regression coefficients

$$\theta_0 + \sum_{j=1}^{J} z_{ij}' \theta_j$$

Group lasso

$$\underset{\theta_0, \theta_j}{\text{argmin}} \left\{ \sum_{i=1}^{n} (y_i - \theta_0 - \sum_{j=1}^{J} z_{ij}' \theta_j)^2 + \lambda \sum_{j=1}^{J} \|\theta_j\|_2 \right\}$$

Properties

- All elements of $\hat{\theta}_j$ will be either zero or non-zero
- With groups of single covariates, group lasso equals lasso