Recap

Adaptive Lasso (Zou, 2006)

two-stage procedure:

- initial estimator $\hat{\beta}_{\text{init}}$, e.g., the Lasso
- re-weighted ℓ_1-penalty

$$\hat{\beta}_{\text{adapt}}(\lambda) = \text{argmin}_{\beta} \left(\|Y - X\beta\|_2^2/n + \lambda \sum_{j=1}^{p} \frac{|\beta_j|}{|\hat{\beta}_{\text{init},j}|} \right)$$

at least as sparse (typically more sparse) than Lasso

$\leadsto$ good/“better” for very sparse underlying mechanisms/models

KKT (Karush-Kuhn-Tucker) conditions

necessary and sufficient conditions for a solution of the Lasso objective function

$$G_j(\hat{\beta}) = -\mathrm{sign}(\hat{\beta}_j)\lambda \text{ if } \beta_j \neq 0$$
$$|G_j(\hat{\beta})| \leq \lambda \text{ if } \beta_j = 0$$

where

$$G(\beta) = -2X^T(Y - X\beta)/n$$

(subdifferential must contain the zero element)

sparsity is potentially induced at points of non-differentiability (here the components of β_j)

Coordinate descent algorithms

for optimization, exploiting the KKT conditions

path following algorithms:
compute $\{\hat{\beta}_j(\lambda)\}_{j=1}^{p}$ over all values of $\lambda \in \mathbb{R}^+$

the coefficient paths are typically “non-monotone” in the non-zeros
it may happen that

$$\hat{\beta}_j(\lambda) \neq 0,\ \hat{\beta}_j(\lambda') = 0 \text{ for } \lambda' < \lambda$$

Generalized Linear Models (GLMs)

univariate response Y, covariate $X \in \mathcal{X} \subseteq \mathbb{R}^p$

GLM: $Y_1, \ldots, Y_n$ independent

$$g(\mathbb{E}[Y_i|X_i = x]) = \underbrace{\mu + \sum_{j=1}^{p} \beta_j x^{(j)}}_{=f(x)=f_{\mu,\beta}(x)}$$

$g(\cdot)$ real-valued, known link function

Lasso: ℓ_1-norm regularized maximum likelihood estimation

$$\hat{\mu}, \hat{\beta} = \text{argmin}_{\mu,\beta}(\underbrace{-\ell(\mu, \beta)}_{\text{neg. log-likelihood}} + \lambda \|\beta\|_1)$$