Recap

Group Lasso (Yuan and Lin, 2006)

groups $\mathcal{G}_1, \ldots, \mathcal{G}_q$ which build a partition of $\{1, \ldots, p\}$
write the (high-dimensional) parameter vector as

$$\beta = (\beta_{\mathcal{G}_1}, \beta_{\mathcal{G}_2}, \ldots, \beta_{\mathcal{G}_q})^T$$

goal: an estimator which is “group-sparse”, i.e.:
for all $j = 1, \ldots, p$,

$$\text{either } \hat{\beta}_{\mathcal{G}_j} \equiv 0$$

$$\text{or } (\hat{\beta}_{\mathcal{G}_j})_r \neq 0 \ \forall r \in \mathcal{G}_j$$

prime example: factor covariates which are encoded by a certain “dummy encoding”
$\rightsquigarrow$ group sparsity then correspond to sparsity in main effects and interaction terms of factor variables

Group Lasso:

$$\hat{\beta}(\lambda) = \text{argmin}_{\beta} \left(\|Y - X\beta\|_2^2/n + \lambda \sum_{j=1}^{q} m_j \|\beta_{\mathcal{G}_j}\|_2 \right)$$

where typically $m_j = \sqrt{|\mathcal{G}_j|}$

group sparsity because objective function is non-differentiable at $\|\beta_{\mathcal{G}_j}\|_2 = 0 \iff \beta_{\mathcal{G}_j} \equiv 0$ $(j = 1, \ldots, q)$

The generalized Group Lasso penalty

$$\text{pen}(\beta) = \lambda \sum_{j=1}^{q} m_j \sqrt{\beta_{\mathcal{G}_j}^T A_j \beta_{\mathcal{G}_j}}, \quad A_j \text{ positive definite}$$

important example: groupwise prediction penalty

$$\text{pen}(\beta) = \sum_{j=1}^{q} m_j \|X_{\mathcal{G}_j} \beta_{\mathcal{G}_j}\|_2 = \lambda \sum_{j=1}^{q} m_j \sqrt{\beta_{\mathcal{G}_j}^T X_{\mathcal{G}_j}^T X_{\mathcal{G}_j} \beta_{\mathcal{G}_j}}$$

$X_{\mathcal{G}_j}^T X_{\mathcal{G}_j}$ typically positive definite for $|\mathcal{G}_j| < n$

- penalty is invariant under arbitrary reparameterizations within every group $\mathcal{G}_j$: important!
- when using an orthogonal parameterization such that $X_{\mathcal{G}_j}^T X_{\mathcal{G}_j} = I$: it is the standard Group Lasso

High-dimensional additive models

$$Y_i = \mu + \sum_{j=1}^{p} f_j(X_i^{(j)}) + \varepsilon_i \ \ (i = 1, \ldots, n;\ p \gg n)$$

$$f_j : \mathbb{R} \to \mathbb{R} \text{ smooth},\ \mathbb{E}[f_j(X_i^{(j)})] \equiv 0\ \forall\ j$$

aim: estimator such that either $\hat{f}_j(.) \equiv 0$ or $\hat{f}_j(.)$ is not the zero function
$\rightsquigarrow$ Group Lasso type problem

for $p < n$: regularize w.r.t. smoothness
for $p \gg n$: need to additionally regularize w.r.t. sparsity