Recap

Group Lasso: $\text{pen}(\beta) = \lambda \sum_{j=1}^{q} m_j \|\beta_{\mathcal{G}_j}\|_2$

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

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

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

- penalty is invariant under arbitrary reparameterizations (bijective mappings) 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

$$Y_i = \mu + \sum_{j=1}^{p} f_j^0(X_i^{(j)}) + \varepsilon_i \;\; (i = 1, \ldots, n;\; p \gg n)$$

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

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

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

Basis expansion and Group Lasso

- basis functions
  $h_{j,k}(.),\ k = 1, \ldots, K$ for every component $j = 1, \ldots, p$
- $n \times K$ matrices

$$H_j :\ (H_j)_{ik} = h_{j,k}(X_i^{(j)})$$

- approximation

$$f_j^0(X_i^{(j)}) \approx \sum_{k=1}^{K} \beta_{j,k} h_{j,k}(X_i^{(j)})$$
$$(f_j^0(X_1^{(j)}), \ldots, f_j(X_n^{(j)}))^T \approx H_j\beta_j,\ \beta_j = (\beta_{j,1}, \ldots, \beta_{j,K})^T$$

Group Lasso with groupwise prediction penalty (and assume $\mu = 0$):

$$\hat{\beta}_1, \ldots \hat{\beta}_p = \text{argmin}_{\beta_1,\ldots,\beta_p} \|Y - \sum_{j=1}^{p} \underbrace{H_j\beta_j}_{=:f_j} \|_2^2 + \lambda \sum_{j=1}^{p} \underbrace{\|H_j\beta_j\|_2/\sqrt{n}}_{=\|f_j\|_n}$$

Sparsity Smoothness (SpS) penalty

$$\hat{f}_1, \ldots, \hat{f}_p = \text{argmin}_{f_1,\ldots,f_p \in \mathcal{F}} \big(\| Y - \sum_{j=1}^{p} f_j \|_n^2 + \lambda_1 \sum_{j=1}^{p} \|f_j\|_n + \lambda_2 \sum_{j=1}^{p} I(f_j) \big)$$

$$\|f_j\|_n^2 = n^{-1} \sum_{i=1}^{n} |f_j(X_i^{(j)})|^2, \; I(f_j)^2 = \int f_j''(x)^2 dx$$

where $\mathcal{F}$ = Sobolev space of functions that are continuously differentiable with square integrable second derivatives

$$\hat{f}_1, \ldots, \hat{f}_p = \text{argmin}_{f_1,\ldots,f_p \in \mathcal{F}} \left(\| Y - \sum_{j=1}^{p} f_j \|_n^2 + \lambda_1 \sum_{j=1}^{p} \| f_j \|_n + \lambda_2 \sum_{j=1}^{p} I(f_j) \right)$$

is a parametric problem of dimension $d \approx 3pn$, parametrized by natural cubic splines with basis functions encoded in a matrix

$$H_{n \times d} = (H_1, \ldots, H_p)^T$$

and integrated squared second derivatives encoded in a matrix

$$(W_j)_{k,\ell} = \int h''_{j,k}(x) h''_{j,\ell}(x) dx$$

$\leadsto$

$$\hat{\beta} = \text{argmin}_{\beta} \left(\| Y - H\beta \|_2^2 / n + \lambda_1 \sum_{j=1}^{p} \sqrt{\beta_j^T H_j^T H_j \beta_j / n} + \lambda_2 \sum_{j=1}^{p} \sqrt{\beta_j^T W_j \beta_j} \right)$$