Recap

Oracle inequality for the Lasso

Theorem 6.1 in Bühlmann and van de Geer (2011)

assume: compatibility condition holds with compatibility constant ϕ_0^2 ($\geq L > 0$)

Then, on $\mathcal{T}$ and for $\lambda \geq 2\lambda_0$:

$$\|X(\hat{\beta} - \beta^0\|_2^2/n + \lambda\|\hat{\beta} - \beta^0\|_1 \leq 4\lambda^2 s_0/\phi_0^2$$

recall: $\mathcal{T} = \{2 \max_{j=1,\dots,p} |\varepsilon^T X^{(j)}/n| \leq \lambda_0\}$

$\mathbb{P}[\mathcal{T}]$ large if $\lambda_0 \asymp \sqrt{\log(p)/n}$

and errors have Gaussian or sub-Gaussian distribution implications:

$$\|X(\hat{\beta} - \beta^0\|_2^2/n = O_P(s_0 \log(p)/n) \quad \text{(fast rate)}$$
$$\|\hat{\beta} - \beta^0\|_1 = O_P(s_o\sqrt{\log(p)/n})$$

these are the (minimax) optimal rates:

no other method can do better

Variable Screening

assume compatibility condition and (e.g.) Gaussian errors
in addition, require beta-min condition:

$$\min_{j \in S_0} |\beta_j^0| \gg s_0 \sqrt{\log(p)/n}$$

$$\Longrightarrow \mathbb{P}[\hat{S} \supseteq S_0] \to 1 \ (p \geq n \to \infty)$$

with high probab: Lasso selects a superset of the active set S_0
$\leadsto$ Lasso does not miss an important active variable!

in practice: $\lambda = \lambda_{CV} \leadsto$ leads "typically" to a too large model

LASSO: Least Absolute Shrinkage and Screening Operator

originally (Tibshirani, 1996):
LASSO: Least Absolute Shrinkage and Selection Operator

Theory versus Practice

theory:

$$\mathbb{P}[\hat{S} \supseteq S_0] \to 1$$

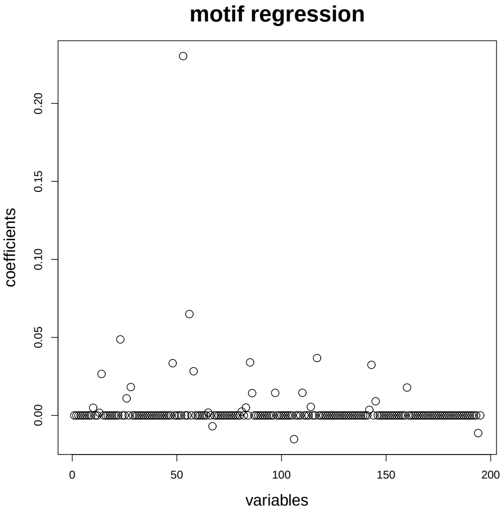
if the following hold:

- compatibility condition for the (fixed) design X
- beta-min condition
- i.i.d. Gaussian errors (can be relaxed)

in addition: $|\hat{S}| \leq \min(n, p)$
hence: huge dimensionality reduction if $p \gg n$

in practice: $\mathbb{P}[\hat{S} \supseteq S_0]$ may not be so large...
$\rightsquigarrow$ see later e.g. “Stability Selection”

The Lasso workhorse



$$p = 195, n = 143, |\hat{S}(\lambda_{CV})| = 26$$

Variable selection

under more restrictive irrepresentable condition or
neighborhood stability condition on the design X
and assuming beta-min condition $\min_{j \in S_0} |\beta_j^0| \gg \sqrt{s_0 \log(p)/n}$:

$$\mathbb{P}[\hat{S} = S_0] \to 1 \ (n \to \infty)$$

the irrepresentable condition is sufficient and essentially necessary for consistent variable selection

this condition is often not fulfilled in practice
(and choosing the correct λ would be difficult as well)

$\leadsto$ variable screening is realistic ("choose λ by CV")
variable selection is not very realistic

LASSO = Least Absolute Shrinkage and Screening Operator

rather than original name, indicating unrealistic picture:
LASSO = Least Absolute Shrinkage and Selection Operator

version of Table 2.2 in the book:

property	design condition	size of non-zero coeff.
slow prediction conv. rate	no requirement	no requirement
fast prediction conv. rate	compatibility	no requirement
estimation error bound $\|\hat{\beta} - \beta^0\|_1$	compatibility	no requirement
variable screening	compatibility or restricted eigenvalue	beta-min condition weaker beta-min cond.
variable selection	neighborhood stability $\Leftrightarrow$ irrepresentable cond.	beta-min condition