Sample questions for exam High-Dimensional Statistics

Peter Bühlmann

December 4, 2019

The set-up

One statement which is either true or false. A correct answer (by checking T or F) will give 1 point, an incorrect answer -1 point (and no answer 0 point).

Lasso for linear models.
Consider a high-dimensional linear model

$$Y = X\beta^0 + \varepsilon,\ \varepsilon \sim \mathcal{N}(0, \sigma^2 I_{n\times n}) \text{ with } p \gg n.$$

We always assume that the $n \times p$ design matrix X is fixed (deterministic).

1. Consider the Lasso estimator $\hat{\beta}(\lambda)$ with regularization parameter λ. If $\hat{\beta}_j(\lambda) \neq 0$ then $\hat{\beta}_j(\lambda') \neq 0$ for all $\lambda' \leq \lambda$.

2. Without assuming anything on the design matrix X but making some sparsity assumptions on β, the Lasso is consistent, saying that

$$X(\hat{\beta} - \beta) \to 0 \ (p \geq n \to \infty) \text{ in probability.}$$

4. Consider the case where the error variables in the linear model satisfy:

$$\varepsilon_1, \ldots, \varepsilon_n \text{ jointly independent with } \mathbb{E}[\varepsilon_i] = 0 \text{ and } \mathrm{Var}(\varepsilon_i) = \sigma_i^2 \leq C_{\mathrm{upp}} < \infty\ \forall i.$$

   This situation is known as “heteroscedastic error structure”.

   All the probabilistic error bounds for the Lasso still hold in such a setting (where instead of the variance σ^2 one can simply plug-in the upper bound C_{upp}).

The de-biased Lasso for linear models.

5. The de-biased Lasso $\hat{b}_1$ achieves a convergence rate for β_1^0 of order $O(1/\sqrt{n})$, asymptotically as both $p \geq n$ tend to infinity (assuming reasonable regularity conditions, e.g. sufficient sparsity of β^0, compatibility condition on the design X and the regression of X_1 versus X_{-1} is sufficiently sparse).