Norm-regularized empirical risk minimization

Sara van de Geer

ETH Zürich

The usefulness of ℓ_1-norm regularization in high-dimensional problems is nowadays well recognized. A fundamental property of the ℓ_1-norm that allows for adaptive estimation and oracle results is its *decomposability*. The ℓ_1-norm of a vector $\beta \in \mathbb{R}^p$ is

$$\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|.$$

With *decomposability* of $\| \cdot \|_1$ we mean that for any set $S \subset \{1, \ldots, p\}$,

$$\|\beta\|_1 = \|\beta_S\|_1 + \|\beta_{-S}\|_1,$$

where $\beta_S = \{\beta_j : \ j \in S\}$ and $\beta_{-S} := \{\beta_j : \ j \notin S\}$. In this talk, we review results for alternative norms Ω on $\mathbb{R}^p$. Fix some $\beta \in \mathbb{R}^p$. We call Ω *weakly decomposable* at β if

$$\Omega \geq \Omega^+ + \Omega^-,$$

where Ω^+ and Ω^- are semi-norms on $\mathbb{R}^p$, and where $\Omega^+(\beta) = \Omega(\beta)$ and $\Omega^-(\beta) = 0$. We will show sharp oracle results - depending on the (approximate) weak decomposability of Ω at certain "oracle" values β - for empirical risk minimizers with regularization penalty proportional to Ω. We also present an approach based on the so-called *triangle property*. We say that Ω has the triangle property at β if there exists semi-norms Ω^+ and Ω^- such that for all β'

$$\max_{z \in \partial\Omega(\beta)} z^T(\beta' - \beta) \geq \Omega^-(\beta') - \Omega^+(\beta' - \beta).$$

Here, $\partial\Omega(\beta)$ is the sub-differential of Ω at β. Several examples with various loss functions (least squares, minus log-likelihood) and penalties (wedge penalty, nuclear norm penalty) will illustrate the theory.