

Sparsity, oracles and inference in high-dimensional statistics

Sara van de Geer

There will be three lectures, which in principle will be independent units. Their common theme is exploiting sparsity in high-dimensional statistics. Sparsity means that the statistical model is allowed to have quite a few parameters, but that it is believed that most of these parameters are actually not relevant. We let the data themselves decide which parameters to keep by applying a regularization method. The aim is then to derive so-called sparsity oracle inequalities.

In the first lecture, we consider a statistical procedure called M-estimation. “M” stands here for “minimum”: one tries to minimize a risk function, in order to obtain the best fit to the data. Least squares is a prominent example. Regularization is done by adding a sparsity inducing penalty that discourages too good a fit to the data. An example is the ℓ_1 -penalty which together with least squares gives to an estimation procedure called the Lasso. We address the question: why does the ℓ_1 -penalty lead to sparsity oracle inequalities and how does this generalize to other norms?

We will see in the first lecture that one needs conditions which relate the penalty to the risk function. They have in a certain sense to be “compatible”. We discuss these compatibility conditions in the second lecture in the context of the Lasso, where the ℓ_1 -penalty needs to be compatible with the least squares risk, i.e. with an ℓ_2 -norm. We give as example the total variation penalty. For $D := \{x_1, \dots, x_n\} \subset \mathbb{R}$ an increasing sequence, the total variation of a function $f : D \rightarrow \mathbb{R}$ is the sum of the absolute values of its jump sizes. We derive compatibility and as a consequence a sparsity oracle inequality which shows adaptation to the number of jumps.

In the third lecture we use sparsity to establish confidence intervals for a parameter of interest. The idea is to use the penalized estimator as an initial estimator in a one-step Newton-Raphson procedure. Functionals of this new estimator that can under certain conditions be shown to be asymptotically normally distributed. We show that in the high-dimensional case, one may further profit from sparsity conditions if the inverse Hessian of the problem is not sparse.