High-dimensional generalized linear models and the LASSO

Sara van de Geer
ETH Zürich

We study a regression model with many more unknown parameters than observations, and show that the LASSO can provide good predictions. Let $X \in \mathcal{X}$ be the dependent variable, and $Y \in \mathbf{R}$ be the response. For a "prediction" $f(X)$ of Y, we consider some loss $\gamma(Y, f(X))$. We let the model class $\mathcal{F}$ be a given p-dimensional linear space of real-valued functions f on $\mathcal{X}$, say

$$\mathcal{F} = \{f_\alpha = \sum_{k=1}^{p} \alpha_k \psi_k(\cdot) : \ \alpha \in \mathbf{R}^p\},$$

with the $\{\psi_k\}$ given functions on $\mathcal{X}$.

Let the observations $\{(X_i, Y_i)\}_{i=1}^n$ consist of n i.i.d. copies of (X, Y). The empirical risk is

$$R_n(f) := \frac{1}{n} \sum_{i=1}^{n} \gamma(Y_i, f(X_i)).$$

When p is large, minimizing $R_n(f)$ over all $f \in \mathcal{F}$ will generally result in overfitting. The LASSO regularizes the empirical risk by adding a penalty proportional to the weighted ℓ_1 norm

$$\hat{I}(\alpha) = \sum_{k=1}^{p} \hat{\tau}_k \alpha_k,$$

with weights $\hat{\tau}_k$ equal to the empirical standard deviation of ψ_k. Taking smoothing parameter $\lambda_n > 0$, the LASSO estimator is

$$\hat{f}_n = \arg \min_{f_\alpha \in \mathcal{F}} \left\{ R_n(f_\alpha) + \lambda_n \hat{I}(\alpha) \right\}.$$

The value of λ_n is chosen of order $\sqrt{\log p / n}$. We will show that under general conditions, in "sparse" situations, the estimator has good predictive properties, in the sense that $R(\hat{f}_n)$ is close to the overall minimum $\min_{\text{all } f} R(f)$. Here, $R(f) := \mathbf{E}\gamma(Y, f(X))$ is the theoretical risk.

The examples include least squares, robust regression, logistic regression and support vector machine type loss.