On high dimensional data analysis in case of no sparsity

Silvelyn Zwanzig

Uppsala University, Sweden, silvelyn.zwanzig@math.uu.se

Keywords: high-dimensional, p larger than n, lasso, ridge

Consider a data set

$$\big(\mathbf{Y},\mathbf{X}\big) = \begin{pmatrix} Y_1 & X_{1,1} & \cdots & X_{p,1} \\ \vdots & \vdots & \ddots & \vdots \\ Y_n & X_{1,n} & \cdots & X_{p,n} \end{pmatrix}. \tag{1}$$

When

$$p > n \ \text{ or even } \ p \gg n,$$

the data are high-dimensional. A formal relationship between $\mathbf{Y}$ and $\mathbf{X}$ to be considered is through the linear model

$$\mathbf{Y} = \mathbf{X}\beta_0 + \epsilon, \tag{2}$$

where $\beta \in \mathbb{R}^p$ is the parameter vector, $\epsilon \in \mathbb{R}^n$ is the error vector of i.i.d. elements assumed to follow certain distribution with $\mathrm{E}(\epsilon) = \mathbf{0}$. The most frequently adopted route to tackle the problem of high-dimensionality is regularization by which a penalty term, $\mathrm{pen}(\beta)$, is added to the least-squares criterion. The regularized least-squares objective function is

$$\min_{\beta\in\mathbb{R}^p} \left[\|\mathbf{Y} - \mathbf{X}\beta\|^2 + \lambda \mathrm{pen}(\beta)\right], \tag{3}$$

where the *tuning parameter* λ controls the intensity of penalization. The ridge estimator uses $\mathbb{L}_2$ norm as the penalty. LASSO combines least-squares L_2 loss with L_1 penalty. There now exist many variants of original LASSO with different penalty terms. In literature the basic idea is to set a condition by which not all covariates are needed, although it is unknown which of them can be deleted. The true parameter $\beta_0 = (\beta_1, \ldots, \beta_p)^T$ satisfies sparsity condition when

$$\|\beta_0\| = \sum_{j=1}^{p} |\beta_j| = o\left(\sqrt{\frac{n}{\log(p)}}\right). \tag{4}$$

In the talk the least squares estimator and different penalized estimators are compared under non-sparsity.