Project Proposal

Benjamin Noland

Background

A common problem in applied statistics is estimation of a vector $\beta^* \in \mathbb{R}^p$ of unknown but fixed parameters in the linear model

$$y = X\beta^* + \epsilon, \tag{1}$$

where $y \in \mathbb{R}^n$ is a vector of observed responses, $X \in \mathbb{R}^{n \times p}$ is the design matrix, and $\epsilon \in \mathbb{R}^n$ is a zero-mean random vector representing the uncertainty in the model.

In the classical setting, we assume that the number of parameters p is small relative to the number of observations, specifically $p \leq n$. In this setting, assuming the design matrix X has full row rank, straightforward linear algebra yields an explicit, unique least-squares estimator of β^*.

However, the situation when there are more parameters than observations, i.e., $p > n$, is not so well understood, and belongs to the active area of research known as *high-dimensional statistics*. One of the strategies commonly employed in high-dimensional statistics is to assume that the data is *truly low-dimensional* in some sense. In the context of our linear model (1), this means assuming that a large number of the entries of the true parameter vector β^* are zero. To be precise, define the *support* of β^* by

$$S(\beta^*) = \{i \in \{1, \ldots, p\} : \beta_i^* \neq 0\},$$

and let $k = |S(\beta^*)|$ denote its cardinality, i.e., the number of non-zero entries of β^*. We assume that the vector β^* is *sparse*, in the sense that $k \ll p$. Under this *sparsity assumption*, the problem reduces to that of computing the support $S(\beta^*)$, allowing us to identify which parameters in the vector β^* are truly important. In this way, we have the potential to substantially reduce the dimensionality of the original problem.

A computational tractable method for computing estimates of the parameters β^* in the high-dimensional setting is the *LASSO* [2] (Least Absolute Shrinkage And Selection Operator). The LASSO computes an estimate of β^* as a solution to the following l_1-constrained quadratic program:

$$\begin{array}{ll} \text{minimize} & \|y - X\beta\|_2^2 \\ \text{subject to} & \|\beta\|_1 \leq C_n \end{array},$$

or equivalently, as the solution to the unconstrained problem

$$\text{minimize } \frac{1}{2n} \left\| y - X\beta \right\|_2^2 + \lambda_n \left\| \beta \right\|_1 ,$$

where $\lambda_n > 0$ is a *regularization parameter* that is in one-to-one correspondence with C_n via Lagrangian duality [1].

Project details

This project will explore the contributions of the paper [1] to the problem of inferring the support $S(\beta^*)$ of β^* (i.e., the problem of *support recovery*) in the linear model (1) using the LASSO as a means of estimating β^*.

Overview of the paper

The paper [1] provides both necessary and sufficient conditions for the LASSO to recover the *signed support* $\mathbb{S}_\pm(\beta^*) \in \mathbb{R}^p$ of β^* with high probability, where $\mathbb{S}_\pm(\beta)$ is defined as follows for any $\beta \in \mathbb{R}^p$:

$$\mathbb{S}_\pm(\beta)_i = \begin{cases} +1 & \text{if } \beta_i > 0 \\ -1 & \text{if } \beta_i < 0 \\ 0 & \text{if } \beta_i = 0 \end{cases} \quad (i = 1, \ldots, p).$$

Specifically, the authors consider the following two questions:

- What relationships between n, p, and k yield a *unique* LASSO solution $\hat{\beta}$ satisfying $\mathbb{S}_\pm(\hat{\beta}) = \mathbb{S}_\pm(\beta^*)$?
- For what relationships between n, p, and k does *no solution* of the LASSO yield the correct signed support?

These questions are analyzed for both deterministic designs and random designs in the linear model (1).

In addition to providing theoretical guarantees, the authors describe the results of simulations to investigate the success/failure of the LASSO in recovering the true signed support for random designs under each of the following sparsity regimes:

- *linear sparsity*: $k(p) = \lceil \alpha p \rceil$ for some $\alpha \in (0, 1)$;
- *sublinear sparsity*: $k(p) = \lceil \alpha p / \log(\alpha p) \rceil$ for some $\alpha \in (0, 1)$, and
- *fractional power sparsity*: $k(p) = \lceil \alpha p^\delta \rceil$ for some $\alpha, \delta \in (0, 1)$.

In each case, the number of observations n is taken to be proportional to $k \log(p-k)$. The true support of the parameter vector is chosen at random.

For each sparsity regime and for several values of p, the authors compute a sequence of values of the *rescaled sample size* $\theta = n/(k\log(p-k))$ and for each such value, compute a sequence of corresponding LASSO solutions $\hat{\beta}$ in order to approximate the probability $\mathrm{P}\{\mathbb{S}_{\pm}(\hat{\beta}) = \mathbb{S}_{\pm}(\beta^*)\}$ of recovering the true signed support. This approximated probability is then plotted against the rescaled sample size θ.

The first round of experiments samples the design matrix $X \in \mathbb{R}^{n\times p}$ from a uniform Gaussian ensemble; that is, its rows are sampled independently from the distribution $N_p(0, I_p)$. A second round of experiments samples X from a non-uniform Gaussian ensemble; specifically, one such that the rows are sampled independently from the distribution $N_p(0, \Sigma)$, where Σ is a $p \times p$ Toeplitz matrix of the form

$$\Sigma = \begin{pmatrix} 1 & \mu & \mu^2 & \cdots & \mu^{p-2} & \mu^{p-1} \\ \mu & 1 & \mu & \mu^2 & \cdots & \mu^{p-2} \\ \mu^2 & \mu & 1 & \mu & \cdots & \mu^{p-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu^{p-1} & \cdots & \mu^3 & \mu^2 & \mu & 1 \end{pmatrix},$$

for some $\mu \in (-1, +1)$. In both cases, the authors note good agreement with the their theoretical predictions.

Expected analysis

In addition to replicating the results of the simulation in [1], I intend to run some simulations in order to predict the probability of support recovery using penalty functions other than l_1 norm. These simulations will be similar in implementation to those from [1] (in particular, the true support to be recovered will be randomly generated). Some possibilities include:

- Ridge regression (i.e., l_2 penalization) with small coefficients thresholded to zero;
- SCAD-penalization [3].

The R package `ncvreg`[1] provides efficient implementations of each of these models.

References

[1] Wainwright, M. (2006). *Sharp thresholds for high-dimensional and noisy sparsity recovery using l_1-constrained quadratic programming (Lasso).* Technical Report 709, Dept. Statistics, Univ. California, Berkeley

[1] `https://cran.r-project.org/web/packages/ncvreg/index.html`

[2] Tibshirani, R. (1996). *Regression shrinkage and selection via the Lasso.* J. Roy. Statist. Soc. Ser. B **58** 267–288

[3] Fan, J. and Li, R. (2001). *Variable selection via nonconcave penalized likelihood and its oracle properties.* J. Amer. Statist. Assoc. **96** 1348–1360