High Dimensional Statistics

Spring 2025

Sharif University of Technology

Dr. Amir Najafi

Homework 5 | Sparse Linear Models in High Dimensions | Due: 1404/4/10

Problem 1: Optimization and Threshold Estimators *[10 points]*

(a) Show that the hard-thresholding estimator (7.6a in the source book) corresponds to the optimal solution $\tilde{\theta}$ of the non-convex program

$$\min_{\theta \in \mathbb{R}^n} \left\{ \frac{1}{2} \|y - \theta\|_2^2 + \frac{1}{2} \lambda^2 \|\theta\|_0 \right\}.$$

(b) Show that the soft-thresholding estimator (7.6b in the source book) corresponds to the optimal solution $\tilde{\theta}$ of the ℓ_1-regularized quadratic program

$$\min_{\theta \in \mathbb{R}^n} \left\{ \frac{1}{2} \|y - \theta\|_2^2 + \lambda \|\theta\|_1 \right\}.$$

Problem 2: Pairwise Incoherence *[15 points]*

Given a matrix $X \in \mathbb{R}^{n \times d}$, suppose that it has pairwise incoherence (7.12) upper bounded as $\delta_{\mathrm{pw}}(X) < \gamma$.

(a) Let $S \subset \{1, 2, \ldots, d\}$ be any subset of size s. Show that there is a function $\gamma \mapsto c(\gamma)$ such that $\gamma_{\min}\left(\frac{X_S^T X_S}{n}\right) \geq c(\gamma) > 0$, as long as γ is sufficiently small.

(b) Prove that X satisfies the restricted nullspace property with respect to S as long as $\gamma < 1/3$. (Do this from first principles, without using any results on restricted isometry.)

Problem 3: Pairwise Incoherence and RIP for Isotropic Ensembles *[10 points]*

Consider a random matrix $X \in \mathbb{R}^{n \times d}$ with i.i.d. $N(0, 1)$ entries.

(a) For a given $s \in \{1, 2, \ldots, d\}$, suppose that $n \gtrsim s^2 \log d$. Show that the pairwise incoherence satisfies the bound $\delta_{\mathrm{pw}}(X) < \frac{1}{3s}$ with high probability.

(b) Now suppose that $n \gtrsim s \log\left(\frac{es}{s}\right)$. Show that the RIP constant satisfies the bound $\delta_s < 1/3$ with high probability.

Problem 4: ℓ_∞-Bounds for the Lasso *[10 points]*

Consider the sparse linear regression model $y = X\theta^* + w$, where $w \sim N(0, \sigma^2 I_{nxn})$ and $\theta^* \in \mathbb{R}^d$ is supported on a subset S. Suppose that the sample covariance matrix $\Sigma = \frac{1}{n}X^T X$ has its diagonal entries uniformly upper bounded by one, and that for some parameter $\gamma > 0$, it also satisfies an ℓ_∞-curvature condition of the form

$$\|\Sigma\Delta\|_\infty \geq \gamma\|\Delta\|_\infty \quad \text{for all } \Delta \in \mathcal{C}_3(S).$$

Show that with the regularization parameter $\lambda_n = 4\sigma\sqrt{\frac{\log d}{n}}$, any Lasso solution satisfies the ℓ_∞-bound

$$\|\hat{\theta} - \theta^*\|_\infty \leq \frac{6\sigma}{\gamma}\sqrt{\frac{\log d}{n}}$$

with high probability.

Problem 5: Online Sparse Regression Regret Bound *[15 points]*

Let $d, T \in \mathbb{N}$, and fix integers k and k_0 satisfying $1 \leq k < k_0 \leq d$. Consider the following (k, k_0, d)-*online sparse regression* setting over T rounds:

- On each round $t = 1, 2, \dots, T$:
  1. The learner selects a subset $S_t \subseteq [d]$ of size $|S_t| = k_0$.
  2. Nature reveals the feature-vector restricted to S_t:
  $$x_t(S_t) \;=\; \big(x_{t,i}\big)_{i \in S_t}, \quad \text{where } x_t \in \mathbb{R}^d \text{ and } \|x_t\|_2 \leq 1.$$
  3. The learner outputs a prediction $\widehat{y}_t \in [-1, 1]$.
  4. Nature then reveals the true label $y_t \in [-1, 1]$, and the learner incurs square loss
  $$\ell_t(\widehat{y}_t) \;=\; \big(\widehat{y}_t - y_t\big)^2.$$
- Define the class of comparator regressors
  $$\mathcal{W}_k = \big\{\, w \in \mathbb{R}^d : \|w\|_2 \leq 1,\ \|w\|_0 \leq k \big\},$$
  where $\|w\|_0$ counts the nonzero coordinates of w. The *regret* of the learner is
  $$\mathrm{Regret}_T \;=\; \sum_{t=1}^{T}\big(\widehat{y}_t - y_t\big)^2 \;-\; \min_{w \in \mathcal{W}_k}\sum_{t=1}^{T}\big(w \cdot x_t - y_t\big)^2.$$

Prove that there exists an (inefficient) algorithm which, for any $k_0 \geq k + 2$, runs in time

$$O\Big(\binom{d}{k}\,k_0\Big) \quad \text{per round,}$$

and guarantees the expected regret bound

$$\mathbb{E}\big[\mathrm{Regret}_T\big] \;=\; O\Big(\tfrac{d^2}{(k_0 - k)^2}\,\sqrt{k \log d\, T}\Big).$$

Problem 6: Error Bounds in Least-Squares Estimation *[15 points]*

Consider the linear model

$$y = X\beta + z,$$

where $y \in \mathbb{R}^n$ is the response vector, $X \in \mathbb{R}^{n\times p}$ is the predictor matrix, $\beta \in \mathbb{R}^p$ is the true weight vector, and $z \in \mathbb{R}^n$ is a noise vector with i.i.d. entries $z_i \sim \mathcal{N}(0, \sigma_z^2)$.
The least-squares estimate is:

$$\beta_{\text{ls}} = \arg\min_{\tilde{\beta}} \|y - X\tilde{\beta}\|_2.$$

Here, $y \in \mathbb{R}^n$ is the observed response vector, $X \in \mathbb{R}^{n\times p}$ is the matrix of predictors, $\beta \in \mathbb{R}^p$ is the true coefficient vector, and $z \in \mathbb{R}^n$ is Gaussian noise with variance σ_z^2. The vector β_{ls} denotes the least-squares estimator. The notation $\|v\|_2$ refers to the Euclidean (or ℓ_2) norm. The quantities $\sigma_{\min}$ and $\sigma_{\max}$ are the smallest and largest singular values of X, respectively. Assume $p \geq n$ and that X is full-rank (i.e., has rank n). Show that, with high probability,

$$\frac{p\sigma_z^2(1-\epsilon)}{\sigma_{\max}^2} \leq \|\beta - \beta_{\text{ls}}\|_2^2 \leq \frac{p\sigma_z^2(1+\epsilon)}{\sigma_{\min}^2},$$

for some small constant $\epsilon > 0$. Provide a sharp bound on the aforementioned probability.

Reading Material

Read the paper *Confidence Intervals and Hypothesis Testing for High-Dimensional Regression* and answer the following problems.

Problem 7: Paper Summarization *[15 points]*

In this problem, you should write a clear and concise overview of the paper in **no more than two pages**. Please ensure that your summary:

- **Highlights the paper's main goals and contributions.** What are the authors aiming to achieve, and why is it important?
- **Describes the key technical ideas and methods.** How do the authors approach the problem? What techniques or frameworks do they employ?
- **Discusses the significance and implications of the results.** Why are these findings impactful, and how might they influence future research or applications?

Above all, your summary should be well-structured, straightforward, and demonstrate a thorough understanding of the paper's content.

Problem 8: Variance Bound for De-Biased LASSO *[10 points]*

Consider a high-dimensional linear regression model given by

$$Y = \mathbf{X}\theta_0 + W,$$

where:

- $\mathbf{X} \in \mathbb{R}^{n\times p}$ is a deterministic design matrix with $p > n$,

- $Y \in \mathbb{R}^n$ is the response vector,
- $\theta_0 \in \mathbb{R}^p$ is the true parameter vector, sparse with $\|\theta_0\|_0 = s_0$,
- $W \sim N(0, \sigma^2 I_n)$ is Gaussian noise,
- $\hat{\Sigma} = \frac{1}{n}\mathbf{X}^T\mathbf{X}$ is the sample covariance matrix.

Let $\hat{\theta}^n$ be the LASSO estimator defined as:

$$\hat{\theta}^n = \arg\min_{\theta \in \mathbb{R}^p} \left\{ \frac{1}{2n}\|Y - \mathbf{X}\theta\|_2^2 + \lambda\|\theta\|_1 \right\},$$

with regularization parameter $\lambda = \sigma\sqrt{\frac{c^2 \log p}{n}}$ for some constant $c > 0$. Define the de-biased estimator:

$$\hat{\theta}^u = \hat{\theta}^n + \frac{1}{n}M\mathbf{X}^T(Y - \mathbf{X}\hat{\theta}^n),$$

where $M \in \mathbb{R}^{p\times p}$ satisfies $|M\hat{\Sigma} - I|_\infty \leq \mu$ for a small $\mu > 0$, and $|\cdot|_\infty$ denotes the entrywise ℓ_∞ norm.
Assume that:

- $\mathbf{X}$ satisfies the compatibility condition for the support $S = \mathrm{supp}(\theta_0)$ with constant $\phi_0 > 0$,
- $\max_i \hat{\Sigma}_{ii} \leq K_0 < \infty$.

Prove that the variance of each component $\hat{\theta}_i^u$ of the de-biased estimator satisfies:

$$\mathrm{Var}(\hat{\theta}_i^u) \leq \frac{\sigma^2}{n} \cdot \frac{1}{(1-\mu)^2\hat{\Sigma}_{ii}},$$

for all $i \in [p] = \{1, \ldots, p\}$, under the given assumptions. Interpret the result in the context of high-dimensional inference.