Homework 3 Due: May 12, 2016
Spring 2016
ECE 598
Information-theoretic methods in high-dimensional statistics
Prof. Yihong Wu

Rules:

- It is mandatory to type your solutions in LaTeX. Email your solution in pdf (file name: `HW3-your name.pdf`) by midnight of the due date to yihongwu@illinois.edu with subject line `Homework XX: your name`.
- Justify your work rigorously. As long as you are able to prove the result or a stronger version, there is no need to follow the hints.

1. (Suboptimality of MLE in high dimensions) Consider the p-dimensional GLM: $X \sim \mathcal{N}(\theta, I_p)$, where θ belongs to the parameter space

$$\Theta = \left\{ \theta \in \mathbb{R}^p : |\theta_1| \leq p^{1/4}, \|\theta_{\setminus 1}\|_2 \leq 2(1 - p^{-1/4}|\theta_1|) \right\}.$$

where $\theta_{\setminus 1} = (\theta_2, \ldots, \theta_p)$. For the mean-square error loss, show that for sufficiently large p,

(a) the minimax risk is bounded:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_\theta[\|\hat{\theta} - \theta\|_2^2] \lesssim 1$$

(b) the worst-case risk of maximal likelihood estimator

$$\theta_{\rm MLE} \triangleq \arg\min_{\tilde{\theta} \in \Theta} \|X - \tilde{\theta}\|_2$$

is unbounded, namely,

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta[\|\hat{\theta}_{\rm MLE} - \theta\|_2^2] \gtrsim \sqrt{p}.$$

(c) (Bonus) Can you construct a cleaner example?

2. (Distribution estimation) Given n independent samples $X_1, \ldots, X_n$ drawn from a distribution P over $[k]$, the goal is to show that the minimax rate for estimating P with respect to the total variation loss is given by:

$$\inf_{\hat{P}} \sup_{P \in \mathcal{M}_k} \mathbb{E}_P[d_{\rm TV}(\hat{P}, P)] \asymp \sqrt{\frac{k-1}{n}} \wedge 1, \qquad k, n \in \mathbb{N}, \tag{1}$$

where $\mathcal{M}_k$ denotes all distributions on $[k]$.

(a) Show that the maximal likelihood estimator $P_{\rm MLE}$ coincides with the empirical distribution.

(b) Show that MLE is rate-optimal, namely, achieving the RHS of (1) within constant factor.

(c) Establish the minimax lower bound via Assouad's lemma.

(d) Establish the minimax lower bound via Fano's inequality + volume method or explicit packing.

(e) (Bonus) Establish the minimax lower bound via mutual information method.

3. (Hard thresholding)

   (a) Show that the solution to the ℓ_0-penalized least squares

   $$\hat{\theta} = \arg\min_{\theta \in \mathbb{R}^p} \|y - \theta\|_2^2 + \lambda \|\theta\|_0$$

   coincides with the hard-thresholding estimator $\hat{\theta}^{\mathrm{HT}}$, where

   $$\hat{\theta}_i^{\mathrm{HT}} = \begin{cases} y_i & |y_i| > \tau \\ 0 & |y_i| \leq \tau. \end{cases}$$

   Find the relation between τ and λ.

   (b) Show that $\hat{\theta}^{\mathrm{HT}}$ is a minimizer for

   $$\hat{\theta} = \arg\min_{\theta : \|y - \theta\|_\infty \leq \tau} \|\theta\|_0.$$

4. (Soft thresholding)

   (a) Show that the solution to the ℓ_1-penalized least squares

   $$\hat{\theta} = \arg\min_{\theta \in \mathbb{R}^p} \|y - \theta\|_2^2 + \lambda \|\theta\|_1$$

   coincides with the soft-thresholding estimator $\hat{\theta}^{\mathrm{ST}}$, where

   $$\hat{\theta}_i^{\mathrm{ST}} = \begin{cases} y_i - \tau & y_i > \tau \\ 0 & |y_i| \leq \tau \\ y_i + \tau & y_i < -\tau \end{cases}$$

   Find the relation between τ and λ.

   (b) Show that

   $$\hat{\theta} = \arg\min_{\theta : \|y - \theta\|_\infty \leq \tau} \|\theta\|_1$$

   also coincides with $\hat{\theta}^{\mathrm{ST}}$.

5. (Sharp minimax rate in sparse denoising) For p-dimensional GLM model $X \sim \mathcal{N}(\theta, I_p)$, we show that minimax risk for denoising 1-sparse vector in high dimension is

   $$\inf_{\hat{\theta}} \sup_{\|\theta\|_0 \leq 1} \mathbb{E}_\theta[\|\hat{\theta} - \theta\|_2^2] = (2 + o(1)) \log p, \qquad p \to \infty. \tag{2}$$

   (a) (Bayesian lower bound) Consider the prior π under which θ is uniformly distributed over $\{\tau e_1, \ldots, \tau e_p\}$, where e_i's denote the standard basis. Let $\tau = \sqrt{(2 - \epsilon) \log p}$. Show that for any $\epsilon > 0$, the Bayes risk (MMSE) is given by

   $$\inf_{\hat{\theta}} \mathbb{E}_{\theta \sim \pi}[\|\hat{\theta} - \theta\|_2^2] = (2 + o(1)) \log p, \qquad p \to \infty.$$

   (Hint: use any method you prefer, e.g., mutual information, compute the exact conditional mean and conditional variance, etc.)

   (b) Demonstrate an estimator $\hat{\theta}$ that achieves the RHS of (2) asymptotically. (Hint: one idea is to use soft thresholding with an appropriately chosen threshold).