Example Sheet 3

Topics in Statistical Theory
Part III, Lent 2016 - Quentin Berthet*

Exercise 1

We recall that $X \in \mathbf{R}^{n\times d}$ is said to satisfy the *restricted isometry property* for sparsity k, with parameter $\alpha \in (0,1)$ if, for all subsets S of $\{1,\dots,d\}$ of cardinality k, we have that

$$1-\alpha \le \|Xv\|_2^2 \le 1+\alpha\,,$$

for all $v \in \mathbf{R}^d$ such that $\|v\|_2 = 1$ with sparsity k.

(a) Show that for all columns X_i of X, we have that $|\|X_i\|_2^2 - 1| \le \alpha$.

(b) Show that if X satisfies the property for sparsity k and parameter α, it satisfies also the property for sparsity $k' \le k$ and parameter $\alpha' \ge \alpha$.

(c) For $n \ge d$, give a sufficient condition on X such that the property is satisfied for some parameter $\alpha \in (0,1)$, for all sparsities $1 \le k \le d$.

(d) For $d > n$, can the property be satisfied for $k > n$ for any $\alpha \in (0,1)$?

(e) A matrix X is drawn by taking independent coefficients X_{ij} such that $\mathbf{E}[X_{ij}] = 0$, $\mathbf{E}[X_{ij}^2] = 1$, and $X_{ij} \in \mathsf{sG}(\sigma^2)$, for some $\sigma > 0$. Show that the matrix $X/\sqrt{n}$ satisfies the property with sparsity k and parameter α with probability at least $1-\delta$ when

$$\alpha \ge C\sqrt{\frac{k\log(d/k) + \log(1/\delta)}{n}}\,.$$

Exercise 2

Let $X_1,\dots,X_n$ be n fixed vectors of $\mathbf{R}^d$. For $1 \le i \le n$, we observe independent (X_i, Y_i), where $Y_i \in \{0,1\}$ has distribution

$$\mathbf{P}(Y_i = 1 \,|\, X_i) = \frac{1}{1+e^{X_i^\top \beta^*}}\,,$$

for some unknown parameter $\beta^* \in \mathcal{C} \subset \mathbf{R}^d$. We wish to estimate β^* based on these observations, a problem called *logistic regression*.

(a) Draw the function $x \mapsto 1/(1+e^x)$, explain how the intuition behind the model.

*q.berthet@statslab.cam.ac.uk

(b) Write the log-likelihood as a function of β. Taking $\sigma_i = 2Y_i - 1$, show that the maximum likelihood estimator can be written as $\hat{\beta} \in \operatorname{argmin}_{\beta \in \mathcal{C}} \varphi_n(\beta)$, where

$$\varphi_n(\beta) = \sum_{i=1}^{n} \log\left(1 + e^{\sigma_i X_i^\top \beta}\right).$$

(c) Show that φ_n is a convex function. Why is this important in practice? Show that if $d > n$ and $\mathcal{C} = \mathbf{R}^d$, the minimizer is not unique.

(d) Compute $\varphi = \mathbf{E}[\varphi_n]$. Show that it is a convex function. Using the notation $\pi_i(\beta) = 1/(1+e^{X_i^\top \beta})$, show that

$$\varphi(\beta) = \varphi(\beta^*) + \sum_{i=1}^{n} \mathsf{KL}(\pi_i(\beta^*), \pi_i(\beta)).$$

Give a minimizer of φ. Give an interpretation of this result.

(e) Show that there exists a random vector $V_n \in \mathbf{R}^d$ such that for all $\beta \in \mathbf{R}^d$, we have

$$\varphi_n(\beta) = \varphi(\beta) + V_n^\top \beta.$$

What can you say about the distribution of V_n?

(f) In the remainder of the problem, let $\|X_i\|_2 \le L$ and $\mathcal{C}$ be a compact set. Show that for some $r_{\mathcal{C},\delta}$ depending on the parameters of the problem, we have with probability $1-\delta$

$$\|\varphi_n - \varphi\|_{\infty,\mathcal{C}} = \max_{\beta \in \mathcal{C}} |\varphi_n(\beta) - \varphi(\beta)| \le r_{\mathcal{C},\delta}.$$

Give an explicit form for $r_{\mathcal{C},\delta}$ when $\mathcal{C} = \mathcal{B}_1^d(R)$, the ℓ_1 ball of radius $R > 0$.

(g) Show that we have with probability at least $1-\delta$

$$\sum_{i=1}^{n} |\pi_i(\beta^*) - \pi_i(\hat{\beta})|^2 \le 2 r_{\mathcal{C},\delta}.$$

(h) Show that for $\mathcal{C} \subset \mathcal{B}_2^d(R)$, we have for some $c_R > 0$.

$$|\pi_i(\beta^*) - \pi_i(\hat{\beta})| \ge c_R |X_i^\top(\hat{\beta} - \beta^*)|.$$

Together with the assumptions above, let $X/\sqrt{n}$ be also *well-conditioned* on $\mathcal{C} - \mathcal{C}$, i.e. that

$$\frac{\|Xv\|_2}{\sqrt{n}} \ge \kappa_{X,\mathcal{C}} \|v\|_2$$

for all $v \in \mathcal{C} - \mathcal{C}$. Show that we have with probability at least $1-\delta$

$$\|\hat{\beta} - \beta^*\|_2^2 \le \frac{1}{n} \, \frac{2\, r_{\delta,\mathcal{C}}}{c_R^2\, \kappa_{X,\mathcal{C}}^2}.$$

Give an explicit bound when $\mathcal{C} = \mathcal{B}_1^d(\rho)$.