Mathematical Machine Learning Theory, M1407.002700
E. Ryu
Spring 2024

Homework 3
Due 5pm, Tuesday, April 9, 2024

Problem 1: *Numerical resolution of LS.* When solving the least squares problem in practice, it is more efficient to use the QR decomposition than to use the SVD. Given $\Phi \in \mathbb{R}^{N\times d}$ with full column rank, the QR factorization has the form

$$\Phi = QR,$$

where $Q \in \mathbb{R}^{N\times d}$ contains orthonormal columns and $R \in \mathbb{R}^{d\times d}$ is upper triangular. Show the following.

(a) R has non-zero diagonal components, i.e., $R_{ii} \neq 0$ for $i = 1, \dots, d$.

(b) Assuming $\Phi = QR$, has already been computed, propose an algorithm for computing $\hat{\theta} = \Phi^{\dagger} Y$ for $Y \in \mathbb{R}^N$. The algorithm may not use a matrix inverse or utilize any matrix decomposition aside from the already computed QR decomposition.

Problem 2: *Linear regression in the random design setting is harder than the fixed design setting.* Consider the least square estimator

$$\hat{\theta} = (\Phi^\intercal \Phi)^{-1} \Phi^\intercal.$$

Recall that the expected excess risk of the least-squares estimator is

$$\mathbb{E}[\mathcal{R}(\hat{\theta})] - \mathcal{R}^\star = \frac{\sigma^2 d}{N}$$

for the fixed design setting, where we assume $\widehat{\Sigma} = \frac{1}{N}\Phi^\intercal \Phi$ is invertible, and

$$\mathbb{E}[\mathcal{R}(\hat{\theta})] - \mathcal{R}^\star = \frac{\sigma^2}{N}\mathbb{E}[\mathrm{Tr}(\Sigma \widehat{\Sigma}^{-1})]$$

for the random design setting, where we assume $\widehat{\Sigma}$ is invertible almost surely and $\Sigma = \mathbb{E}_X[\phi(X)\phi(X)^\intercal]$ is invertible. Show that

$$\frac{\sigma^2}{N}\mathbb{E}[\mathrm{Tr}(\Sigma \widehat{\Sigma}^{-1})] \geq \frac{\sigma^2 d}{N}.$$

You may use the following fact without proof: The mapping $M \mapsto \mathrm{Tr}(M^{-1})$ is convex on the set of symmetric positive definite matrices. Do not assume $\phi(X_1)$ is Gaussian.

Hint. Define $Z = \Phi \Sigma^{-1/2}$ as in the lecture, and use Jensen on $\mathbb{E}[\mathrm{Tr}(Z^\intercal Z)^{-1}]$.

Problem 3: *Convex functions have convex sublevel sets.* Let $f\colon \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$. Define the *$\alpha$-sublevel set* of f as

$$C_\alpha = \{x \in \mathbb{R}^d \mid f(x) \le \alpha\}.$$

Show that if f is convex as a function, then C_α is convex as a set.

Problem 4: *Convex functions have convex epigraphs.* Let $f\colon \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$. Define the *epigraph* of f as

$$\mathrm{epi}(f) = \{(x,t) \mid f(x) \le t,\, x \in \mathbb{R}^d,\, t \in \mathbb{R}\} \subset \mathbb{R}^{d+1}.$$

Show that f is convex as a function if and only if $\mathrm{epi}(f)$ is convex as a set.

Problem 5: *Convexity of maximum eigenvalues.* Show that $\lambda_{\max}$, as a function on the set of symmetric matrices, is convex.

Hint. Use $\lambda_{\max}(M) = \sup_{\|v\|=1} v^\intercal M v$.

Problem 6: *Projection onto convex sets is well defined.* Let $A \subseteq \mathbb{R}^d$ be a nonempty closed convex set and let $p \in \mathbb{R}^d$. Show that

$$\operatorname*{argmin}_{x \in A} \|x - p\|_2,$$

where $\|\cdot\|_2$ denotes the Euclidean norm, exists and is unique.

Hint. For uniqueness, show that $f(x) = \|x-p\|^2$ is *strictly convex* function and then argue that if $x, x' \in A$ are two distinct minimizes, then $\frac{1}{2}x + \frac{1}{2}x' \in A$ would be closer to p.

Problem 7: *A subgradient may not exist on the boundary of the domain.* Let $f\colon [0,\infty) \to \mathbb{R}$ defined by $f(x) = -\sqrt{x}$. Show that f does not have a subgradient at $x = 0$, i.e., there is no g such that

$$f(y) \ge f(0) + g \cdot y, \qquad \forall\, y \in [0,\infty).$$

Problem 8: *A subgradient provides a cutting plane for* $\operatorname{argmin} f$. Let $f\colon \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ be convex. Show the following.

(a) If $g \in \partial f(x)$ and $g \neq 0$, then

$$\operatorname{argmin} f \subset \{y \in \mathbb{R}^d \mid g^\intercal y \le g^\intercal x\}.$$

(b) If f is differentiable, and $g = \nabla f(x) \neq 0$, then

$$\operatorname{argmin} f \subset \{y \in \mathbb{R}^d \mid g^\intercal y < g^\intercal x\}.$$

Problem 9: *Closure of convex set is convex.* Let $C \subseteq \mathbb{R}^d$ be a convex set, and let $\overline{C} \subseteq \mathbb{R}^d$ be its closure. (Closure in the sense of open sets and closed sets.) Show that $\overline{C}$ is convex.

Problem 10: *Strict separating hyperplane theorem.* Let $C \subseteq \mathbb{R}^d$ be a nonempty open convex set. Let $p \in \mathbb{R}^d$ be such that $p \notin C$. Then, there is a non-zero $v \in \mathbb{R}^d$ such that

$$v^\intercal x < v^\intercal p, \quad \forall\, x \in C.$$

Hint. Consider the two cases $p \notin \partial C$ and $p \in \partial C$ and work with $\overline{C}$.

Problem 11: *Expectation on a convex set is in the convex set.* Let $C \subseteq \mathbb{R}^d$ be a nonempty open convex set. Let $X \in \mathbb{R}^d$ be a random variable such that $X \in C$ almost surely and $\mathbb{E}[X] \in \mathbb{R}^d$ is well defined. Show that $\mathbb{E}[X] \in C$.

Hint. Assume for contradiction that $\mathbb{E}[X] \notin C$. Then there is a strict separating hyperplane between $\mathbb{E}[X]$ and C given by v. Consider $\mathbb{E}[v^\intercal X]$.

Remark. The statement holds even if C is a nonempty convex set (not necessarily open). The proof of the general case involves extending the arguments of this exercise using the notion of relative interiors.

Problem 12: *Jensen for φ with open convex domain.* Let $C \subseteq \mathbb{R}^d$ be a nonempty open convex set. Let $X \in \mathbb{R}^d$ be a random variable such that $X \in C$ almost surely and $\mathbb{E}[X] \in \mathbb{R}^d$ is well defined. Let $\varphi\colon C \to \mathbb{R}$ be convex. Show that

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

Remark. Jensen's inequality holds even if C is a nonempty convex set (not necessarily open). The proof of the general case involves extending the arguments of this exercise using the notion relative interiors.