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

Homework 4
Due 5pm, Thursday, April 25, 2024

Problem 1: *Gaussian expectations for matrix Bernstein.* Let $Z \sim \mathcal{N}(0, I_{d\times d})$. Show that

$$\lambda_{\max}\left(\mathbb{E}[\|Z\|^2 ZZ^\intercal]\right) = 2 + d.$$

Hint. Note that if $Z = [Z_1, \dots, Z_d]^\intercal$, then $\mathbb{E}[Z_1] = 0$, $\mathbb{E}[Z_1^2] = 1$, $\mathbb{E}[Z_1^3] = 0$, $\mathbb{E}[Z_1^4] = 3$, and

$$\left(\mathbb{E}[\|Z\|^2 ZZ^\intercal]\right)_{ij} = \sum_{k=1}^{d} \mathbb{E}[Z_k^2 Z_i Z_j], \qquad \text{for } i, j \in \{1, \dots, d\}.$$

Problem 2: *Exercise with Pseudo-inverse.* Let $\Phi^{N\times d}$ and let $\Phi^\dagger$ be the pseudo-inverse of Φ. Show that

$$\Phi^\intercal(\Phi\Phi^\dagger - I) = 0.$$

Problem 3: *Solution to ridge regression.* Let $\Phi \in \mathbb{R}^{N\times d}$ and $\widehat{\Sigma} = \frac{1}{N}\Phi^\intercal\Phi \in \mathbb{R}^{d\times d}$. Let $\mu > 0$. Consider the optimization problem

$$\underset{\theta\in\mathbb{R}^d}{\text{minimize}} \quad \tfrac{1}{N}\|Y - \Phi\theta\|^2 + \mu\|\theta\|_2^2.$$

Do not make any assumptions about the rank of Φ or $\widehat{\Sigma}$.

(a) Show that

$$\hat{\theta}_\mu = \frac{1}{N}(\widehat{\Sigma} + \mu I)^{-1}\Phi^\intercal Y = (\Phi^\intercal\Phi + N\mu I)^{-1}\Phi^\intercal Y = \Phi^\intercal(\Phi\Phi^\intercal + N\mu I)^{-1} Y$$

is the unique solution to the optimization problem.

(b) Generically, if $A \in \mathbb{R}^{\ell\times m}$ and $B \in \mathbb{R}^{m\times n}$, then computing AB requires $\mathcal{O}(\ell mn)$ computation. If $C \in \mathbb{R}^{n\times n}$ is invertible, computing C^{-1} requires $\mathcal{O}(n^3)$ computation. What are the computational costs of directly computing $(\Phi^\intercal\Phi + N\mu I)^{-1}\Phi^\intercal Y$ and $\Phi^\intercal(\Phi\Phi^\intercal + N\mu I)^{-1} Y$?

Hint. For (a), use the matrix inversion lemma.

Problem 4: *Cocoercivity inequality from L-smoothness.* Let $F\colon \mathbb{R}^d \to \mathbb{R}$ be L-smooth convex with $0 < L < \infty$. Show that

$$\langle \nabla F(\theta) - \nabla F(\eta), \theta - \eta \rangle \geq \frac{1}{L}\|\nabla F(\theta) - \nabla F(\eta)\|^2, \qquad \forall\, \theta, \eta \in \mathbb{R}^d.$$

Problem 5: *Strong monotonicity inequality from μ-convexity.* Let $F\colon \mathbb{R}^d \to \mathbb{R}$ be differentiable and μ-strongly convex with $0 < \mu < \infty$. Show that

$$\langle \nabla F(\theta) - \nabla F(\eta), \theta - \eta \rangle \geq \mu\|\theta - \eta\|^2, \qquad \forall\, \theta, \eta \in \mathbb{R}^d.$$

Problem 6: *Distance to solution ⇒ function-value suboptimality.* Let $0 < L < \infty$. Let $F\colon \mathbb{R}^d \to \mathbb{R}$ be L-smooth convex with minimizer $\theta^\star$. Assume we have shown a guarantee of

$$\|\theta^k - \theta^\star\|^2 \leq h(k)\|\theta^0 - \theta^\star\|^2$$

for $k = 0, 1, \ldots$, where $h\colon \mathbb{N} \to [0, \infty)$. Show that

$$F(\theta^k) - F(\theta^\star) \leq \frac{Lh(k)}{2}\|\theta^0 - \theta^\star\|^2$$

for $k = 0, 1, \ldots$.

Problem 7: *Quadratic objectives only need symmetric matrices.* Let

$$F(\theta) = \theta^\intercal H\theta + b^\intercal\theta + c,$$

where $H \in \mathbb{R}^{d\times d}$, $b \in \mathbb{R}^d$, and $c \in \mathbb{R}$. Show that

$$F(\theta) = \frac{1}{2}\theta^\intercal(H + H^\intercal)\theta + b^\intercal\theta + c.$$

Problem 8: *Convex quadratic objectives.* Let

$$F(\theta) = \frac{1}{2}\theta^\intercal H\theta + b^\intercal\theta + c,$$

where $H = H^\intercal \in \mathbb{R}^{d\times d}$, $b \in \mathbb{R}^d$, and $c \in \mathbb{R}$.

(a) Show that $F(\theta)$ is convex if and only if $H \succeq 0$.

(b) Show that if H has a negative eigenvalue, then $\inf_{\theta\in\mathbb{R}^d} F(\theta) = -\infty$.

Problem 9: *Quadratic objectives in standard form.* Let

$$F(\theta) = \frac{1}{2}\theta^\intercal H\theta + b^\intercal\theta + c,$$

where $H = H^\intercal \in \mathbb{R}^{d\times d}$ is strictly positive definite, $b \in \mathbb{R}^d$, and $c \in \mathbb{R}$. Show that there exists some $\theta^\star \in \mathbb{R}^d$ and $c' \in \mathbb{R}$ such that

$$F(\theta) = \frac{1}{2}(\theta - \theta^\star)^\intercal H(\theta - \theta^\star) + c'.$$