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

Homework 6
Due 5pm, Friday, May 31, 2024

Problem 1: *Expectation of convex functions.* Let $\omega \sim P$ be a random variable. Let $f(\cdot;\omega) \colon \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ be convex and assume $f(\cdot;\omega) \ge 0$ for (P-almost) all ω. Let

$$F(x) = \mathop{\mathbb{E}}_{\omega \sim P}[f(x;\omega)], \forall\, x \in \mathbb{R}^d.$$

(a) Show that $F(x)$ is well defined and $F \colon \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$, i.e., $F(x)$ is never $-\infty$.

(b) Show that F is convex.

Remark. To be measure-theoretically precise, assume $f(x;\cdot)$ is P-measurable for all $x \in \mathbb{R}^d$.

Problem 2: *Inactive constraints can be dropped in convex optimization.* Let $f \colon \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ and $g \colon \mathbb{R}^d \to \mathbb{R}$ be CCP. Let $x^\star$ be a solution to the optimization problem

$$\begin{array}{ll} \underset{x \in \mathbb{R}^d}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \le D. \end{array}$$

Assume $g(x^\star) < D$. Show that $x^\star$ is a solution to

$$\underset{x \in \mathbb{R}^d}{\text{minimize}} \quad f(x).$$

Remark. The implication is not true without convexity.

Problem 3: *Constrained to regularized formulation.* Let $f \colon \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ and $g \colon \mathbb{R}^d \to \mathbb{R}$ be CCP. Let $x^\star$ be a solution to the optimization problem

$$\begin{array}{ll} \underset{x \in \mathbb{R}^d}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \le D. \end{array}$$

Assume the strict feasibility condition: there exists an $\tilde{x} \in \mathrm{dom} f$ such that $g(\tilde{x}) < D$. Then there is a $\lambda \ge 0$ such that $x^\star$ is a solution to

$$\underset{x \in \mathbb{R}^d}{\text{minimize}} \quad f(x) + \lambda g(x).$$

Remark. The lecture provided a reasonably complete outline of the proof. This problem is asking you to fill in the gaps.

Problem 4: *Analysis of SGD with smoothness.* Let $0 < L < \infty$. Consider the stochastic optimization problem

$$\underset{x\in\mathbb{R}^d}{\text{minimize}} \quad \mathbb{E}_{\omega}[f(x;\omega)] = F(x),$$

where ω is a random variable. Assume $F\colon \mathbb{R}^d \to \mathbb{R}$ is convex L-smooth, and assume F has a minimizer $x^\star$. Consider stochastic gradient descent with constant stepsize

$$x^{k+1} = x^k - \alpha g^k$$

for $k = 0, 1, \ldots$, where g^k is a stochastic gradient of F at x^k. Assume the stochastic gradient g^k satisfies

$$\mathbb{E}_k[g^k] = \nabla F(x^k), \qquad \mathrm{Var}_k(g^k) = \mathbb{E}_k[\|g^k - \nabla F(x^k)\|^2] \le \sigma^2$$

for $k = 0, 1, \ldots$. Let $x^0 \in \mathbb{R}^d$ be a starting point. Let $K > 0$ be the total iteration count. Consider SGD with the constant stepsize

$$\alpha_k = \alpha = \frac{\|x^0 - x^\star\|_2}{\sigma\sqrt{K+1}}.$$

(a) For $k = 0, 1, \ldots, K$, show

$$\mathbb{E}\big[\|x^{k+1} - x^\star\|^2 \,\big|\, x^0, \ldots, x^k\big] \le \|x^k - x^\star\|^2 - 2\alpha\big(F(x^k) - F(x^\star)\big) - \alpha\big(\tfrac{1}{L} - \alpha\big)\|\nabla F(x^k)\|^2 + \alpha^2\sigma^2.$$

(b) Assume K is large enough so that $\alpha \le 1/L$. Show

$$\mathbb{E}\big[f(\bar{x}^K;\omega) - f(x^\star;\omega)\big] = \mathbb{E}[F(\bar{x}^K) - F(x^\star)] \le \frac{\sigma\|x^0 - x^\star\|_2}{\sqrt{K+1}},$$

where

$$\bar{x}^K = \frac{1}{K+1}\sum_{k=0}^{K} x^k.$$

Problem 5: Let $\mathcal{X} = (-1, 1)$. Show that $K\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ defined as

$$K(x, x') = \frac{1}{1 - xx'}$$

is PDK.

Problem 6: *Basic exercise on PDK.* Let $\mathcal{Z}$ be a nonempty set and let

$$\mathcal{X} = \{A \subseteq \mathcal{Z} \,|\, |A| < \infty\}.$$

Show that $K\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ defined as

$$K(A, A') = 2^{|A \cap A'|}$$

is a PDK.

Clarification. $|A|$ denotes the cardinality of the set A.
Hint. Consider

$$\sum_{S \subseteq A \cup A'} \mathbf{1}_{S \subseteq A} \mathbf{1}_{S \subseteq A'}.$$

Problem 7: *Basic exercise on PDK.* Let $\mathcal{Z}$ be a topological space, let μ be a nonnegative finite Borel measure on $\mathcal{Z}$, and let

$$\mathcal{X} = \{A \subseteq \mathcal{Z} \,|\, A \text{ is Borel measurable}\}.$$

Show that $K\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ defined as

$$K(A, B) = \mu(A \cap B)$$

is a PDK.

Hint. Consider the feature map $\phi(A) = \mathbf{1}_A \in L^2(\mu)$.

Problem 8: *Simple RKHS facts.* Let $\mathcal{H}$ be an RKHS of functions defined on $\mathcal{X}$ with RK $K\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Show the following.

(i) $K(x, x) \geq 0$ for all $x \in \mathcal{X}$.

(ii) If $f_k \to f_\infty$ in $\mathcal{H}$, then $f_k(x) \to f_\infty(x)$ for all $x \in \mathcal{X}$.

(iii) Define $d_K\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ as

$$d_K(x, x') = \|K(\cdot, x) - K(\cdot, x')\|_{\mathcal{H}}.$$

Then d_K is a pseudometric on $\mathcal{X}$. If K further is strictly positive definite, then d_K is a metric on $\mathcal{X}$.

(iv) If $K(x, x) = 0$, then $K(x, x') = 0$ for all $x' \in \mathcal{X}$.

(v) The *normalized kernel* $\tilde{K}\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ defined as

$$\tilde{K}(x, x') = \begin{cases} \frac{K(x,x')}{\sqrt{K(x,x)K(x',x')}} & \text{if } K(x,x)K(x',x') > 0 \\ 0 & \text{otherwise} \end{cases}$$

is a PDK.