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

Homework 1
Due 5pm, Wednesday, March 13, 2024

Problem 1: *Norms are convex.* Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$. Show that $\|\cdot\|$ is convex.

Problem 2: *Dual norms are norms.* Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$. Show that the dual norm $\|\cdot\|_*$ defined as

$$\|y\|_* = \sup\{y^\intercal x \,|\, x \in \mathbb{R}^d,\ \|x\| \le 1\}$$

for all $y \in \mathbb{R}^d$ is a norm on $\mathbb{R}^d$.

Problem 3: *Bound on surrogate loss bounds original loss.* Consider the binary classification problem, where $(X, Y) \sim P$ with $X \in \mathbb{R}^d$ and $Y \in \{-1, +1\}$. Let $f(x) = \operatorname{sign}(g(x))$, where $g\colon \mathbb{R}^d \to \mathbb{R}$ and $\operatorname{sign}(0) = 0$. Let

$$\mathcal{R}_{\Phi^{\text{logistic}}}[g] = \mathbb{E}_{(X,Y)\sim P}[\Phi^{\text{logistic}}(Yg(X))], \qquad \Phi^{\text{logistic}}(u) = \log(1 + e^{-u}).$$

Show that

$$\mathcal{R}_{\Phi^{\text{logistic}}}[g] < \varepsilon \quad \Rightarrow \quad \mathbb{P}(f(X) = Y) > 1 - \frac{\varepsilon}{\log 2}$$

for any $\varepsilon > 0$.

Problem 4: *Practice with Jensen and indicator functions.* Let $X \in \mathbb{R}$ be a random variable. Show the following.

(i) If $\mathbb{E}[X] \in \mathbb{R}$ is well defined, i.e., $\mathbb{E}[|X|] < \infty$, then

$$\mathbb{E}[X] \le \mathbb{E}[|X|] \le \sqrt{\mathbb{E}[X^2]}.$$

(ii) It always holds, even when $\mathbb{E}[|X|] = \infty$, that

$$\mathbb{E}[|X|] \le \sqrt{\mathbb{E}[X^2]}.$$

Hint. Use Jensen. For the case $\mathbb{E}[|X|] = \infty$, use

$$|X| = |X|\mathbf{1}_{\{|X|\le 1\}} + |X|\mathbf{1}_{\{|X|>1\}}, \qquad X^2 = X^2\mathbf{1}_{\{|X|\le 1\}} + X^2\mathbf{1}_{\{|X|>1\}}.$$

Problem 5: *Max of expectation $\le$ expectation of max.* Let $f\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. Let P be a probability measure on $\mathcal{X}$. Show that

$$\sup_{y\in\mathcal{Y}} \mathbb{E}_{X\sim P}[f(X, y)] \le \mathbb{E}_{X\sim P}[\sup_{y\in\mathcal{Y}} f(X, y)].$$

(Assume $\sup_{y\in\mathcal{Y}} f(x, y)$ is P-measurable.)

Problem 6: *Equality in distribution.* Let $X \in \mathcal{X}$ and $Y \in \mathcal{X}$ be random variables. We say X and Y are *equal in distribution* and write $X \stackrel{\mathcal{D}}{=} Y$ if and only if

$$\mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(Y)]$$

for all bounded $\varphi \colon \mathcal{X} \to \mathbb{R}$. Let us furthermore assume that $X \in \mathbb{R}$ and $Y \in \mathbb{R}$ are continuous random variables with density functions f_X and f_Y. Show that $X \stackrel{\mathcal{D}}{=} Y$ if and only if $f_X = f_Y$.

Remark. More generally, it is possible to show (using simple functions) that $X \stackrel{\mathcal{D}}{=} Y$ if and only if the probability measures for $X \in \mathcal{X}$ and $Y \in \mathcal{X}$ are equal.

Problem 7: *Equal almost surely vs. in distribution.* We say $X \in \mathcal{X}$ and $Y \in \mathcal{X}$ are equal (almost surely) and write $X = Y$ if the outcomes of X and Y are the same with probability 1.

(a) Construct a pair of random variables X and Y such that $X \stackrel{\mathcal{D}}{=} Y$ but $X \neq Y$.

(b) Show that $X = Y$ implies $X \stackrel{\mathcal{D}}{=} Y$.

Problem 8: *Symmetric random variables.* We say a random variable $X \in \mathbb{R}^d$ is *symmetric* if $X \stackrel{\mathcal{D}}{=} -X$. Let $X_1, \ldots, X_N$ be scalar-valued independent symmetric random variables. Let $\varepsilon_1, \ldots, \varepsilon_N$ be IID Rademacher random variables, i.e., $\varepsilon_i \in \{-1, +1\}$ with probability 1/2 for each outcome for $i = 1, \ldots, N$. Assume $X_1, \ldots, X_N, \varepsilon_1, \ldots, \varepsilon_N$ are mutually independent. Show

$$f(X_1, \ldots, X_N) \stackrel{\mathcal{D}}{=} f(\varepsilon_1 X_1, \ldots, \varepsilon_N X_N)$$

for any f.

Hint. Show that

$$\mathbb{E}[\varphi(f(X_1, \ldots, X_N))] = \mathbb{E}[\varphi(f(\varepsilon_1 X_1, \ldots, \varepsilon_N X_N))]$$

for any bounded φ by conditioning on $\varepsilon_1, \ldots, \varepsilon_N$, i.e., first take the conditional expectation of $X_1, \ldots, X_N$ given $\varepsilon_1, \ldots, \varepsilon_N$.

Problem 9: *Expected error to PAC.* Let $\mathcal{R}^\star = \inf_g \mathcal{R}[g]$ so that $\mathcal{R}[g] - \mathcal{R}^\star \geq 0$. Assume

$$\mathbb{E}_g \left[\mathcal{R}[g] - \mathcal{R}^\star\right] < \delta$$

holds for some $\delta > 0$. Show that

$$\mathbb{P}_g \left(\mathcal{R}[g] - \mathcal{R}^\star < \varepsilon\right) > 1 - \frac{\delta}{\varepsilon}, \qquad \forall \varepsilon > 0.$$

Problem 10: *Generalized Markov.* Let $\varphi : [0, \infty) \to [0, \infty)$ be a strictly increasing function and $X \in \mathbb{R}$ a nonnegative random variable. Show that

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{1}{\varphi(\varepsilon)} \mathbb{E}[\varphi(X)], \qquad \forall \varepsilon > 0.$$

Problem 11: Let $X_1, \ldots, X_N \in \mathbb{R}$ be independent sub-Gaussian random variables with constant τ. Show that $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$ is sub-Gaussian with constant $\tau/\sqrt{N}$.

Problem 12: Let $X \in [a, b]$ be a continuous random variable with density function $f(x)$. Show that X is sub-Gaussian with constant $\tau^2 = (b - a)^2/4$.