CS7545, Fall 2018: Machine Learning Theory - Homework #1

Jacob Abernethy, Bhuvesh Kumar, Jun-Kun Wang Due: Monday, September 10 at 2 p.m.

Homework Policy: Homeworks must be typed up in LaTeX and submitted as a pdf. Working in groups is fine, but you must write up your own solutions. Please write the members of your group on your solutions. There is no strict limit to the size of the group but we may find it a bit suspicious if there are more than 4 to a team. Questions labeled with **(Challenge)** are not strictly required, but you'll get some participation credit if you make some meaningful progress on the question.

1) **Norm.** Prove the following norm inequalities. Assume $\mathbf{x} \in \mathbb{R}^N$.

(a) $\|\mathbf{x}\|_2 \le \|\mathbf{x}\|_1 \le \sqrt{N}\|\mathbf{x}\|_2$

(b) $\|\mathbf{x}\|_\infty \le \|\mathbf{x}\|_2 \le \sqrt{N}\|\mathbf{x}\|_\infty$

(c) $\|\mathbf{x}\|_\infty \le \|\mathbf{x}\|_1 \le N\|\mathbf{x}\|_\infty$

(d) $\|\mathbf{x}\|_p \le N^{1/p}\|\mathbf{x}\|_\infty$ for $p > 1$

(e) $\lim_{p\to+\infty} \|\mathbf{x}\|_p = \|\mathbf{x}\|_\infty$ for $p > 1$

2) **Hölder.** Let $\mathbf{p} \in \Delta_N$ with full support; that is, $\sum_{i=1}^N p_i = 1$ and $p_i > 0$ for all $i = 1, \dots, N$.

(a) Prove using Hölder's Inequality that $\sum_{i=1}^N \frac{1}{{p_i}^{q-1}} \ge N^q$ for any $q > 1$.

(b) Prove that $\sum_{i=1}^N \left(p_i + \frac{1}{p_i}\right)^2 \ge N^3 + 2N + 1/N$

3) **Projection.** Given a set $\mathcal{K} \subseteq \mathbb{R}^d$, we define the projection operator $\Pi_{\mathcal{K}}$ as follows for any $x \in \mathbb{R}^d$:

$$\Pi_{\mathcal{K}}(x) = \arg\min_{y\in\mathcal{K}} \|x - y\|_2 \tag{1}$$

That is, $\Pi_{\mathcal{K}}(x)$ is the set of closest points in $\mathcal{K}$ to x.

(a) Let $\mathcal{K} \subseteq \mathbb{R}^d$ be a closed and bounded set. Prove that if $\mathcal{K}$ is convex, then the projection $\Pi_{\mathcal{K}}(x)$ is a singleton (i.e. $|\Pi_{\mathcal{K}}(x)| = 1$) for all $x \in \mathbb{R}^d$.

(b) (**Challenge**) Prove that $\mathcal{K}$ is convex if and only if the projection $\Pi_{\mathcal{K}}(x)$ is a singleton (i.e. $|\Pi_{\mathcal{K}}(x)| = 1$) for all $x \in \mathbb{R}^d$.

(c) If $\mathcal{K} = \{x : \|x\|_2 \leq 1\}$, then for $x \notin \mathcal{K}$, show that $\Pi_{\mathcal{K}}(x) = \frac{x}{\|x\|_2}$. (You may assume statement of problem 3a if neded)

4) **Fenchel.**

(a) Let f, g be convex functions such that $f^* = g$. Write the convex conjugate of $f_\alpha(\cdot) = \alpha f(\cdot)$ (where $\alpha \in \mathbb{R}^+$) in terms of α and g.

(b) Let $f(x) := \sqrt{1 + x^2}$. What is its Fenchel conjugate, $f^*(\theta)$?

5) **Hoeffding.**

(a) **(Challenge)** Let X be a random variable and define $f(\lambda) := \log \mathbb{E}[\exp(\lambda X)]$. Let $f^*(\theta)$ be the Fenchel conjugate of f. Show that for all t such that $t \geq \mathbb{E}[X]$ we have $\mathbb{P}(X \geq t) \leq \exp(-f^*(t))$.

(b) Assume you have m coins. All the coins are *unbiased* (that is, they have an equal probability of heads or tails), EXCEPT the special coin which comes up heads with probability $\frac{1}{2} + \rho$, for some $0 < \rho < \frac{1}{2}$. You toss each of the coins n times exactly, and you count the number of heads you observe; say h_i is the number of times coin i came up heads. Did the special coin have the most heads? Find a lower bound, in terms of n, m, and ρ, on the probability that the special coin had the largest value h_i.

(c) I want to make sure I find the special coin, and I can only handle a small $\delta > 0$ probability of error. Find a value of n, in terms of ρ, δ, and m, to guarantee that the special coin has the most heads with probability at least $1 - \delta$.

6) **Shannon.** Let $f : \Delta_n \to \mathbb{R}$ be the *negative entropy* function, defined as

$$f(\mathbf{x}) = \sum_{i=1}^{n} x_i \log x_i.$$

where $0 \log(0) = 0$.

(a) Let $g : \mathbb{R}^n \to \mathbb{R}$ be a function defined as $g(\boldsymbol{\theta}) = \log\left(\sum_{i=1}^n \exp(\theta_i)\right)$. Show that f is the Fenchel conjugate of g.

(b) Prove that when $n = 2$, f is 1-strongly convex with respect to $\|\cdot\|_1$. Denote two arbitrary vectors in Δ_2 as $\mathbf{x} = (p, 1-p)$ and $\mathbf{y} = (q, 1-q)$. *Hint:* Without loss of generality, assume $p \geq q$.

(c) **(Challenge)** Prove that for any $n \geq 2$, the function f is 1-strongly convex with respect to $\|\cdot\|_1$. *Hint:* Can you find a reduction to the $n = 2$ case? Here is one possible route: Let $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ be two vectors in Δ_n. Let $A = \{i : x_i \geq y_i\}$ be the coordinates where $\mathbf{x}$ dominates $\mathbf{y}$. Find new vectors $\mathbf{x}_A, \mathbf{y}_A \in \Delta_2$ such that $\|\mathbf{x} - \mathbf{y}\|_1 = \|\mathbf{x}_A - \mathbf{y}_A\|_1$ and operate on these.

7) **Bregman.** Let f be a differentiable convex function.

(a) Show that for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \text{dom}(f)$,

$$D_f(\mathbf{x}, \mathbf{y}) + D_f(\mathbf{y}, \mathbf{z}) - D_f(\mathbf{x}, \mathbf{z}) = \langle \nabla f(\mathbf{z}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle.$$

(b) Show that if f is 1-strongly convex with respect to a norm $\|\cdot\|$ then

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_* \geq \|\mathbf{x} - \mathbf{y}\|$$

for all $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$, where $\|\cdot\|_*$ is the dual norm to $\|\cdot\|$. *Hint:* Consider $D_f(\mathbf{x}, \mathbf{y})$ and $D_f(\mathbf{y}, \mathbf{x})$.

(c) **(Challenge)** Show the converse of the above (problem 7b). *Hint:* Let $h(\alpha) = f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}))$ and $\mathbf{z}_\alpha = \mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})$.