# Theoretical Foundations of Machine Learning - Homework #1

Your Name (uniqname), Collaborators: Person 1, Person 2, First Name Suffices — Due: 9/28/2015

**Homework Policy:** Working in groups is fine. 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 labelled with **(Challenge)** are not strictly required, but you'll get some participation credit if you have something interesting to add, even if it's only a partial answer.

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

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

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

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

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

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

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

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

3) **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 $\alpha$ and $g$.

**(b)** Let $f(x) := \sqrt{1 + x^2}$. What is its Fenchel conjugate, $f^*(\theta)$? (EDIT: original version was $f(x) := \sqrt{1 - x^2}$, which is concave!)

4) **Hoeffding.**

**(a)** 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 $\mathbb{P}(X > t) \leq \exp(-f^*(t))$

**(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 probabiliy $\frac{1}{2} + \rho$, 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 $\rho$, 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 $\rho, \delta$, and $m$, to gaurantee that the special coin has the most heads with probability at least $1 - \delta$.

5) **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 $\Delta_2$ 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 $\Delta_n$. 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.

6) **Bregman.** Let $f$ be a differentiable convex function.

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

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

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

$$\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq \|\mathbf{x} - \mathbf{y}\|^2$$

for all $\mathbf{x}, \mathbf{y} \in \mathrm{dom}(f)$, where $\|\cdot\|_*$ is the dual norm to $\|\cdot\|$. *Hint:* Consider $D_f(\mathbf{x}, \mathbf{y})$ and $D_f(\mathbf{y}, \mathbf{x})$. **Note:** The original version asked you to show that $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_* \geq \|\mathbf{x} - \mathbf{y}\|$ which, indeed, is also true. We will accept an answer to either question, but the Challenge question 6(c) refers to the present version.

**(c)** **(Challenge)** Show the converse of the above (problem 6b). *Hint:* Let $h(\alpha) = f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}))$ and $\mathbf{z}_\alpha = \mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})$ .