
CS7545, Spring 2024: Machine Learning Theory - Homework #1

Jacob Abernethy, Tyler Labonte, Guanghui Wang, Yeojoon Youn

Due: Friday, February 2 at 11:59 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.

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 $p > 1$

(e) $\lim_{p \rightarrow +\infty} \|\mathbf{x}\|_p = \|\mathbf{x}\|_\infty$ 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 $\sum_{i=1}^N \frac{1}{p_i^{q-1}} \geq N^q$ 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) **Convexity.**

(a) Show that the following function is convex on the domain $p \in \Delta_N$:

$$f(p) = \sum_{i=1}^N p_i \log p_i.$$

(b) Show that if a differentiable function g is convex then it satisfies the gradient monotonicity property:

$$(\nabla g(x) - \nabla g(y))^T (x - y) \geq 0 \text{ for all } x, y \in \text{dom}(g)$$

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.** Often Hoeffding's Inequality is stated in a different way. Make sure to use Hoeffding to prove this version.

Let X_1, \dots, X_m be m independent random variables sampled from the same distribution D , where D has support on $[-1, 1]$, and the mean of D is μ . Then for some $\alpha, \beta, \gamma > 0$ we have the following statement: with probability at least $1 - \delta$

$$\left| \frac{1}{m} \sum_{i=1}^m X_i - \mu \right| \leq \alpha \sqrt{\frac{\log(\beta/\delta)}{\gamma m}}.$$

When you solve this problem, make sure to get the best values of α, β, γ !

6) **Bayes classifier.** In this problem we will study an example of the Bayes classifier in a binary classification setting where $\mathcal{X} = \mathbb{R}$. Let $Y = 1$ with probability $1/2$ and $Y = -1$ otherwise. Then, we draw X as follows: if $Y = 1$ let $X \sim \mathcal{N}(\mu/2, \sigma^2)$ and if $Y = -1$ let $X \sim \mathcal{N}(-\mu/2, \sigma^2)$ where $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian distribution with mean μ and variance σ^2 .

- (a) Recall that $\eta(x) = \Pr[Y = 1|X = x]$. Show that

$$\eta(x) = \frac{1}{1 + \exp\left(\frac{-x\mu}{\sigma^2}\right)}.$$

****Hint.**** Use Bayes' rule.

- (b) Compute an analytical expression for h_{Bayes} (*i.e.*, substitute $\eta(x)$ and simplify the resulting expression).
- (c) Compute the Bayes error $L(h_{\text{Bayes}})$. You can leave your answer in terms of the Gaussian cdf Φ .
- (d) On which point(s) $x \in \mathbb{R}$ is h_{Bayes} most likely to make a mistake? Why?