# CMPUT 654: Theoretical Foundations of Machine Learning, Fall 2023
# Midterm

## Instructions

**Submissions** You need to submit a single PDF file, named `midterm_<name>.pdf` where `<name>` is your name. The PDF file should include your typed up solutions (we strongly encourage to use pdfLaTeX). Write your name in the title of your PDF file. We provide a LaTeXtemplate that you are encouraged to use. To submit your PDF file you should send the PDF file via private message to Csaba on Slack before the deadline.

**Collaboration and sources** Work on your own. You can use online resources, but no human help can be used. In your write-up you must acknowledge all the sources (books, webpages etc., including class notes.) Failure to do so will be considered cheating. Identical or similar write-ups will be considered cheating. Students are expected to understand and explain all the steps of their proofs.

**Scheduling** Start early: It takes time to solve the problems, as well as to write down the solutions. Most problems should have a short solution (and you can refer to results we have learned about to shorten your solution). Don't repeat calculations that we did in the class unnecessarily.

**Deadline:** October 15, 2023 11:55pm

## Problems

There is only one problem, with many parts. The problem is just to finish the problems in Question 2 of Assignment 1 that were worth zero marks in that Assignments. Good news: Here you will get some good marks for solving these! Also, the last problem (dealing with the absolute value loss) is broken up into many small parts. For the details of the setup, see Assignment 1.

The first question there was asking you to replace the Bretagnolle-Huber inequality with the Neyman-Pearson lemma. The form of Neyman-Pearson lemma that is most useful for our purposes looks as follows:

**Theorem 1** (Neyman-Pearson Lemma). *Let $P$ and $Q$ be probability measures on the same measurable space $(\Omega, \mathcal{F})$, and let $A \in \mathcal{F}$ be an arbitrary event. Let $\mu$ be a measure on $(\Omega, \mathcal{F})$ such that $dP = pd\mu$ and $dQ = qd\mu$. Let $B \in \mathcal{F}$, $\eta \geq 0$ be such that*

*(a)* $P(A) \leq P(B)$,

*(b)* $B \cap \{p > 0 \text{ or } q > 0\} \subset \{q \geq \eta p\}$ *and*

*(c)* $B^c \cap \{p > 0 \text{ or } q > 0\} \subset \{q \leq \eta p\}$.

*Then,*

$$Q(A^c) \geq Q(B^c) \, .$$

This result implies the following corollary, which takes a form very close to the Bretagnolle-Huber inequality:

**Corollary 1.** *Let $A_\eta = \{q \geq \eta p\}$. Let $r > 0$ be such that $(0, r] \subseteq \{P(A_\eta) : \eta \geq 0\}$. Then, either $P(A) \geq r$, or*

$$P(A) + Q(A^c) \geq \inf_{\eta \geq 0} P(A_\eta) + Q(A_\eta^c) \, .$$

Sometimes it is simpler to use a lower bound based on the total variation distance,

$$\delta(P,Q) = \sup_{A\in\mathcal{F}} P(A) - Q(A) (= \sup_{A\in\mathcal{F}} Q(A) - P(A))\,.$$

For this, we have the following result:

**Proposition 1.** *Let $P$ and $Q$ be probability measures on the same measurable space $(\Omega, \mathcal{F})$, and let $A \in \mathcal{F}$ be an arbitrary event. Then,*

$$P(A) + Q(A^c) \geq 1 - \delta(P,Q)\,.$$

**Question 1.**

(Q1) Prove Theorem 1.

**Hint**: Investigate the sign of $(\chi_B - \chi_A)(q - \eta p)$ on $\{p > 0 \text{ or } q > 0\}$. Here, $\chi_A$ ($\chi_B$) is the characteristic function of set $A$ (respectively, $B$): $\chi_A(\omega) = \mathbb{I}\{\omega \in A\}$, $\omega \in \Omega$.

**10 points**

(Q2) Prove Corollary 1

**10 points**

(Q3) Prove Proposition 1

**5 points**

(Q4) Solve Part (i), Question 2: "Use the Neyman-Pearson lemma instead of Theorem 1 to prove an analogue of Equation (6). Compare with Equation(6); the numerical constant in the lower bound should be at least three times larger than there."

**Hint**: Use Corollary 1 together with the inequality

$$\overline{\Phi}(y) \geq \sqrt{\frac{2}{\pi}} \frac{\exp(-y^2/2)}{y + \sqrt{y^2+4}}\,,$$

which holds for any $y \in \mathbb{R}$. Here, $\overline{\Phi}$ denotes the complementary CDF of the standard normal Gaussian: $\overline{\Phi}(z) = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-x^2/2} dx$, $z \in \mathbb{R}$.

**30 points**

(Q5) Solve Part (j), Question 2: "How would the results look like if the quadratic loss was replaced with its scaled version: $\ell_c(u,v) = c(u-v)^2$, $u, v \in \mathbb{R}$, $c > 0$."

**10 points**

(Q6) In the remaining problems we work on solving Part (k), Question 2: "Recalculate the minimax lower bound from Part (g) for the absolute value loss: $\ell(u,v) = |u - v|$, $u, v \in \mathbb{R}$. Also show that $R^*(\sigma^2) \leq \sigma$. What differences do you observe compared to the squared loss?" This is broken up into multiple parts. As it turns out for this loss the median plays the role that the mean played in the case of squared loss. As such, we start with a warm-up problem:

Recall that the median of a distribution $P$ defined over the reals is any value $m \in \mathbb{R}$, such that $\mathbb{P}(Y \le m) \le 1/2$ and $\mathbb{P}(Y \ge m) \le 1/2$, where $Y \sim P$. Fix $P$ in an arbitrary fashion. Let $\mathrm{Med}(P)$ be the set of medians of $P$. Show that $\mathrm{Med}(P)$ is a closed interval (and hence is non-empty).

**10 points**

(Q7) By overloading $L$, introduce $L(P,a) = \int |y-a| P(dy)$ for $a \in \mathbb{R}$. Let $m$ be any median of $P$. Show that $L(P,m) \le L(P,a)$ for any $a \in \mathbb{R}$.

**10 points**

(Q8) Show that $L^*(P) = L(P,m)$.

**5 points**

(Q9) Show that $R^*(\sigma^2) \le \sigma$.

**10 points**

(Q10) We need to turn to the lower bound. Fix $P$ so that zero is a median of $P$ and let

$$r_P(a) = \mathbb{E}[|a-V| - |V|]$$

where $V \sim P$. Show that for any method $\mathcal{A} : \mathbb{R} \to \mathbb{R}$,

$$R(P,\mathcal{A}) = \int r_P(\mathcal{A}(y)) P(dy)\,. \tag{1}$$

**10 points**

(Q11) Show that $r_P$ is differentiable and its derivative, $r'_P$ satisfies $r'_P(a) = P((-\infty,a]) - P([a,\infty))$ for any $a \in \mathbb{R}$ such that $P(\{a\}) \neq 0$.

**Hint**: Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at some point $a \in \mathbb{R}$ with derivative $d \in \mathbb{R}$ if $f(a+h) - f(a) = dh + o(h)$ as $h \to 0$.

**10 points**

(Q12) Let $P$ be a non-atomic distribution; i.e., $P$ is such that for any $a \in \mathbb{R}$, $P(\{a\}) = 0$. Show that $r_P$ is increasing on $[m,\infty)$ and decreasing on $(-\infty,m]$ for any median $m$ of $P$.

**10 points**

(Q13) Let $P$ be any non-atomic distribution, $m \in \mathrm{Med}(P)$ and fix an arbitrary method $\mathcal{A} : \mathbb{R} \to \mathbb{R}$. Show that for any $s \ge 0$,

$$R(P,\mathcal{A}) \ge \max(P(E_+(s)), P(E_-(s)))r^*_P(s)\,, \tag{2}$$

where $r^*_P(s) = \min(r_P(m+s), r_P(m-s))$ and where $E_+(s) = \{y \in \mathbb{R} : \mathcal{A}(y) \ge m+s\}$ and $E_-(s) = \{y \in \mathbb{R} : \mathcal{A}(y) \le m-s\}$.

**5 points**

(Q14) Let $P = \mathrm{Lap}(0, \lambda)$ for $\lambda > 0$ fixed. Show that for any $a \in \mathbb{R}$, $r_P(a) = r_P(-a)$, while for $a \geq 0$,

$$r_P(a) = a - \lambda + \lambda e^{-a/\lambda}\,.$$

**Hint**: Recall that the density function of $\mathrm{Lap}(\mu, \lambda)$, the Laplace distribution with mean (and median) $\mu$ and scale parameter $\lambda$ is

$$p(x) = \frac{1}{2\lambda} e^{-|x-\mu|/\lambda}\,, \qquad x \in \mathbb{R}\,.$$

**5 points**

(Q15) Show that there exist an absolute constant $c > 0$ such that $R^*(\sigma^2) \geq c\sigma$.

**Hint**: Use Laplace distributions for tractability, together with Proposition 1. For using the total variation distance recall that if $dP = p d\mu$ and $dQ = q d\mu$ for some measure $\mu$ on $(\Omega, \mathcal{F})$ then

$$\delta(P, Q) = \frac{1}{2} \int |p - q| d\mu. \tag{3}$$

**15 points**

Total: **155 points**

---

**Total for all questions: 155**. Of this, 30 are bonus marks. Your midterm will be marked out of 125.