CS 281A - Homework 5

November 10, 2025

This assignment is due on **November 21, 2025 at 11:59PM**. Submit your solutions as a **single PDF** on bCourses. You are strongly encouraged to typeset your submission. Illegible submissions will not be graded.

Problem 1

Consider a medical decision making setting with a population of patients all suffering from the same health condition. The condition, left untreated, causes each individual patient i cost $c_i(0)$. A drug for the condition is available, but has known side effects quantified by cost $c_i(1)$, and unknown and probabilistic individual treatment effect quantified by $\tau_i \in \mathbb{R}$, drawn from some unknown distribution with mean τ. If patient i is left untreated, they suffer cost $c_i(0)$; on the other hand, if they are treated with the drug, they suffer cost $c_i(1) - \tau_i$.

1. Suppose only the population average treatment effect $\tau := \mathbb{E}[\tau_i]$ is known. What is the treatment decision rule that minimizes expected cost for a new patient j?

2. Suppose that members of the population are parametrized by features $X \in \{-1, 1\}^d$, and in addition to the population average treatment effect τ, the conditional average treatment effects $\tau_x = \mathbb{E}[\tau_i \mid X_i = x]$ for all x are known. What is the decision rule that minimizes expected cost for a new person j with feature vector x_j?

3. Suppose instead that we make a *prediction* for individual j in the form of an interval $(\widehat{\tau}_j^{\min}, \widehat{\tau}_j^{\max}) \subset \mathbb{R}$, with the guarantee that

$$\mathbb{P}\left[\tau_j \notin (\widehat{\tau}_j^{\min}, \widehat{\tau}_j^{\max})\right] \leq \varepsilon.$$

   Is it possible to design a treatment decision rule for patient j that minimizes the expected cost for j? If so, give such a rule. If not, explain why and design a rule that achieves a different notion of optimality.

4. In words, describe the meaning of optimality for the following scenarios. That is, state the source(s) of randomness, and the degree to which a guarantee holds for a *particular* person j.

   (a) Scenario described in Part 1.

   (b) Scenario described in Part 2.

   (c) Scenario described in Part 3, where the marginal distribution of τ_i over the entire population is known, and the prediction interval $(\widehat{\tau}_j^{\min}, \widehat{\tau}_j^{\max})$ is constructed with this information about the distribution.

   (d) Scenario described in Part 3, where the prediction interval $(\widehat{\tau}_j^{\min}, \widehat{\tau}_j^{\max})$ is constructed by post-processing a point predictor for τ_j with the standard split conformal prediction procedure.

Problem 2

Consider a randomized trial with n units. Let $a_i \in \{0, 1\}$ denote the action taken on unit $i \in [n]$. Let $y_i(j) \in \{0, 1\}$ denote the potential outcome for unit i when the action is j. Recall that the average treatment effect (ATE) is the quantity

$$\text{ATE} = \frac{1}{n} \sum_{i=1}^{n} y_i(1) - y_i(0).$$

1. Suppose n is even. Assume that half of the units are selected uniformly at random and assigned action 1, and the other half are assigned action 0. Prove that the Horvitz–Thompson estimator,

$$\widehat{\mathrm{ATE}} = \frac{2}{n} \sum_{\{i \,:\, a_i = 1\}} y_i(1) - \frac{2}{n} \sum_{\{j \,:\, a_j = 0\}} y_j(0),$$

is unbiased. What is its variance?

2. Now consider a different estimator of the ATE. Suppose m_1 of the units are assigned to action 1 and m_0 to action zero. For $j \in \{0, 1\}$, define

$$\hat{\mu}_j := \frac{1}{m_j} \sum_{\{i \,:\, a_i = j\}} y_i(j)$$

The *z-score* is the quantity

$$z = \frac{\hat{\mu}_1 - \hat{\mu}_0}{\mathrm{std}(y)\sqrt{\frac{1}{m_0} + \frac{1}{m_1}}}.$$

Under the z-test, a result is statistically significant if z has magnitude larger than 1.96. Show that no matter how a is chosen, z is proportional to the correlation coefficient of a and y. That is, prove

$$z = \sqrt{n} \cdot R(a, y) = \frac{\sqrt{n} \cdot \mathrm{cov}(a, y)}{\mathrm{std}(a)\,\mathrm{std}(y)}.$$

Problem 3

Consider a k-armed bandit problem with $[0, 1]$-bounded rewards, which means $X^{(t)} \in [0, 1]^k$ for all rounds t. Let $\mu_i = \mathbb{E}[X_i^{(1)}]$ be the expected reward from arm $i \in [k]$ and let $\mu_\star = \max_{i \in [k]} \mu_i$.

The ε-greedy algorithm takes as input an exploration schedule $\{\varepsilon_t\}$, where $\varepsilon_t \in [0, 1]$ for all t, and works as follows. On the first k rounds, the algorithm plays each arm once. On each round $t > k$, with probability ε_t, the algorithm plays $i_t \sim \mathrm{Unif}(1, \ldots, k)$, and with probability $1 - \varepsilon_t$, it plays

$$i_t = \arg\max_{i \in [k]} \frac{1}{\sum_{s<t} \mathbb{1}(i_s = i)} \sum_{s<t} X_i^{(s)} \cdot \mathbb{1}(i_s = i).$$

1. Suppose $\varepsilon_t = \varepsilon \in (0, 1)$ for all t. Show that

$$\lim_{t \to \infty} \frac{1}{t} \cdot \left(t\mu_\star - \mathbb{E}\left[\sum_{s \le t} X_{i_s}^{(s)} \right] \right) = \frac{\varepsilon}{k} \sum_{i \in [k]} (\mu_\star - \mu_i).$$

What does this result imply about the expected regret of the ε-greedy algorithm with any constant exploration schedule?

2. Suppose $\varepsilon_t = \left(\frac{K \log(t)}{t} \right)^{1/3}$. Show that there is a universal constant $C > 0$ such that

$$t\mu_\star - \mathbb{E}\left[\sum_{s \le t} X_{i_s}^{(s)} \right] \le C \cdot (K \log(t))^{1/3} \cdot t^{2/3}$$

for all t. What does this result imply about the expected regret of the ε-greedy algorithm with a decaying exploration schedule? Compare and contrast this result with the result of the previous part.

Problem 4

Minimizing regret to the best fixed arm in hindsight is only one way to formalize what it means to perform well in a k-armed bandit problem. An alternative goal is *best-arm identification*: at the end of T rounds, predict the index y_T of the arm with highest expected reward. The goal of best-arm identification can be equivalently stated as minimizing instantaneous regret $\mu_\star - \mu_{y_T}$.

When a k-armed bandit problem is interpreted as a clinical trial involving k treatments, the regret minimization problem can be interpreted as minimizing harm done to study participants, while best-arm identification can be interpreted as identifying the best treatment for all future patients, possibly at the cost of greater harm done to study participants.

1. Suppose you alternate between playing each of the k arms over T rounds, such that each arm is played $\lfloor T/k \rfloor$ times. Prove that for sufficiently large T, there exists constant $C > 0$ such that

   $$\mu_\star - \mathbb{E}[\mu_{y_T}] \leq e^{-C \cdot T}.$$

   Can this algorithm achieve sublinear regret as well?

2. Suppose you have a blackbox algorithm whose regret to the best fixed arm in hindsight after T rounds is bounded by a sublinear function of T. That is, the algorithm guaranteees that

   $$T\mu_\star - \mathbb{E}\left[\sum_{t \leq T} X_{i_t}^{(t)}\right] \leq \phi(T)$$

   for some $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{T \to \infty} \phi(T)/T = 0$. Show how this algorithm can be used to obtain the best-arm identification guarantee

   $$\mu_\star - \mathbb{E}[\mu_{y_T}] \leq \frac{\phi(T)}{T}.$$

   What does this result imply about the compatibility of regret minimization and best-arm identification, as goals? Compare and contrast with the result of the previous part.

Problem 5

Please submit the final version of your project abstract.