CS 281A - Homework 2

September 16, 2025

This assignment is due on **September 26, 2025 at 11:59PM**. Submission method TBD.

Problem 1

Let Ω be a set of N bits, each equal to either 0 or 1. For a fixed integer $k < N$, sample a sequence $y_1, \ldots, y_{k+1}$ from Ω uniformly without replacement. Define the sample average of the first k elements as

$$\hat{m}_k = \frac{1}{k} \sum_{i=1}^{k} y_i \,.$$

and define the average of all of the bits as

$$m_N = \frac{1}{N} \sum_{y \in \Omega} y \,.$$

Show that

$$\mathbb{E}[(\hat{m}_k - y_{k+1})^2] = \frac{N}{N-1} \cdot \frac{k+1}{k} \cdot m_N(1 - m_N)\,.$$

Problem 2

Let X be a continuous random variable distributed over the closed interval [0,1]. For label $Y = 0$, X is uniform:

$$p_{X|Y}(X \mid Y = 0) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

For label $Y = 1$, the conditional pdf of X is as follows:

$$p_{X|Y}(X \mid Y = 1) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

1. Find the decision rule that minimizes the probability of error.
2. Find the closed form expression for the operating characteristic of the likelihood ratio test (LRT), i.e., TPR as a function of FPR for the LRT.
3. Suppose we require the TPR to be at least $(1+\epsilon)$FPR, where $\epsilon > 0$ is a fixed constant. Find $\text{TPR}^{\max}(\epsilon)$, the maximal value of TPR that is achievable under this constraint.

Problem 3

Suppose you have n expert forecasters. Each individual forecaster $i \in [n]$ has a known true positive rate and false positive rate $(\text{TPR}_i, \text{FPR}_i)$. Define forecaster 0 to be the artificial forecaster that only predicts $y = 0$ and forecaster $n+1$ to be the artificial forecaster who only predicts $y = 1$.

Consider the set of forecasts defined by the following procedure:

1. Pick two numbers i and j between 0 and $n+1$ (inclusive) and a scalar $p \in [0,1]$.
2. Generate a random number r from the uniform distribution on $[0,1]$.
3. If $r < p$, return the prediction of forecaster i. Otherwise, return the prediction of forecaster j.

1. For a given $t \in [0,1]$, show how to apply this procedure to achieve a TPR of t.
2. What are the possible $(\mathrm{TPR}, \mathrm{FPR})$ pairs you can achieve with this family of forecasts?

Problem 4

In this problem, we consider an automated resume screening tool which is used by a company to sort candidates based on whether or not they are predicted to be invited for an on site interview after an initial phone screen. Let the random variable X denote the features of a candidate's application and Y denote whether a candidate is invited for an on site interview, where $Y = 1$ indicates that an individual was invited.

1. Suppose that there are many qualified individuals looking for jobs and that paying recruiters to call applicants is expensive. As a result, it is comparatively half as costly for the company to miss a candidate who would have been invited on site than it is to spend time calling an individual who is not invited for an interview (i.e. for some $\alpha > 0$, $C_{10} = \alpha$, $C_{01} = \frac{\alpha}{2}$, and other costs are zero). **Show that the company's optimal decision rule for resume screening has the form**

$$s(x) = \mathbb{E}[Y|X = x] \geq t \,,$$

   and find the value of t.

2. Now suppose that unemployment has gone down, and there are no longer many qualified candidates looking for jobs. As a result, it is instead twice as costly to miss good candidates than it is to call ones who are not invited for an interview (i.e. for some $\beta > 0$, $C_{10} = \beta$, $C_{01} = 2\beta$, and other costs are zero). **How does the optimal decision rule change?**

Suppose now that some score function $\hat{s}(x)$ has been estimated from historical data, and a threshold rule is applied to assign individuals the screening predictions $\widehat{Y} = 1$ for those who will be considered more closely by recruiters and $\widehat{Y} = 0$ for those who will not. In the United States, it is illegal to discriminate against job applicants on the basis of religion, and your job is to evaluate this tool with that in mind. Below is a table which shows the predictions and outcomes for applicants split by membership in a minority religious group, with $A = 1$ indicating that an individual is a member of this group and $A = 0$ indicating that they are not. We have data from 500 candidates in the religious group and 5,000 candidates not in the religious group.

	$A = 1$			$A = 0$		
	$Y = 0$	$Y = 1$		$Y = 0$	$Y = 1$	
$\widehat{Y} = 0$	360	40	400	4050	450	4500
$\widehat{Y} = 1$	40	60	100	200	300	500
	400	100		4250	750	

3. With membership in the religious group as the sensitive attribute, **does this classifier satisfy independence? Does it satisfy sufficiency?** Justify your answer.

4. For the criteria that the classifier doesn't satisfy, **propose a group-dependent change to the threshold** that results in a classifier that does satisfy the criteria. You do not need to specify exact quantities, rather comparisons with the current threshold. You should not propose a trivial threshold that results in 0% or 100% acceptance rates.

5. **Compare and contrast** the value of the intervention you suggested in part 4 for the following two circumstances:

- You learn that the historical data comes from a hiring manager who is a member of the religious group and has been heard telling fellow members that they have an "in" regardless of their qualifications.
- You learn that there is a well regarded religious university nearby that sends the resumes of highly qualified students to the company. Historically, these candidates have highly relevant skill sets and make up a majority of applications from the religious group.

Problem 5

Apply the concepts from the course lectures to your final project.

1. Provide your project abstract. This may have changed from last week, and that's ok.
2. What are the features X and labels Y? What could they represent? How are they represented?
3. Describe the sources of these data. Are they simulated, generated, or retrospective? How costly are they to acquire?
4. What are the relevant prediction problems involved with your project? Why are these prediction targets of interest to yourself, to the research community, and to the broader public?
5. Why do you believe that the labels can be predicted from the features?
6. What metrics might you use to evaluate the quality of your predictions? Do you expect there to be tradeoffs between these associated metrics?