Homework #1
CMSC 35401: The Interplay of Economics and Machine Learning (Winter'24)

Due Saturday 01/20, 9:00 pm

General Instructions The assignment is meant to practice your understanding of course materials, and some of them are challenging. You are allowed to discuss with fellow students, however please write up your solutions independently (e.g., start writing solutions after a few hours of any discussion) and, equally importantly, acknowledge everyone you discussed the homework with on your writeup. All course materials are available on the course webiste here https://www.haifeng-xu.com/cmsc35401win24. You may refer to any materials covered in our class. However, any attempt to consult outside sources, on the Internet or otherwise, for solutions to any of these homework problems is *not* allowed.

Whenever a question asks you to "show" or "prove" a claim, please provide a formal mathematical proof. These problems have been labeled based on their difficulties. `Short` problems are intended to take you 5-15 minutes each and `medium` problems are intended to take 15-30 minutes each. `Long` problems may take anywhere between 30 minutes to several hours depending on whether inspiration strikes. Note that, the total score is meant to *not* be normalized to 100 (for instance, this HW has 30 in total for regular students and additional 15 points for those who take it as elective).

Finally, please write your solutions in latex — **hand written solutions will not be accepted.** Hope you enjoy the homework!

Problem 1: Some Facts about Linear Programs

Consider a Linear Program (LP) in the following standard form where $c \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$.

$$\begin{array}{ll} \text{maximize} & c^T \cdot x \\ \text{subject to} & Ax \leq b \\ & x \geq 0 \end{array} \tag{1}$$

Prove the following facts about the LP.

1. (`Short, 4 points`) At any vertex of the feasible region, n linearly independent constraints are satisfied with equality (a.k.a. *tight*). Note that there are $n + m$ linear constraints in LP (1) because $x \geq 0$ account for n linear constraints.

2. (`Short,4 points`) We learned in class that the dual of LP (1) is the following LP (2)

$$\begin{array}{ll} \text{minimize} & b^T \cdot y \\ \text{subject to} & A^T y \geq c \\ & y \geq 0 \end{array} \tag{2}$$

Prove that the dual of LP (2) is the original LP (1).

Problem 2

Prove the following projection lemma and separating hyperplane theorem.

1. (**Projection Lemma**, `Medium, 7 points`) Let $Z \subset \mathbb{R}^n$ be a nonempty closed convex set and $y \notin Z$ be any point in $\mathbb{R}^n$. Prove that there exists $z^* \in Z$ that has the minimum l_2 distance from y among all $z \in Z$. Moreover, $\forall z \in Z$ we have $(y - z^*)^T \cdot (z - z^*) \leq 0$. (**hint**: use Weierstrass' Theorem).

2. (**Separating Hyperplane Theorem**, `Short, 5 points`) Let $Z \subset \mathbb{R}^n$ be a nonempty closed convex set and let $y \notin Z$ be any point in $\mathbb{R}^n$. Prove that there exists a hyperplane $\alpha^T \cdot x = \beta$ that strictly separates y from Z. That is, $\alpha^T \cdot z \geq \beta$ for any $z \in Z$ but $\alpha^T \cdot y < \beta$.

Problem 3: Linear Programming for Machine Learning

In this question, you will learn to formulate some machine learning problems as linear programs. Let us assume that there are n data points $(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)$ where $\mathbf{x}_i \in \mathbb{R}^m$ is interpreted as an m-dimensional feature vector and $y_i \in \mathbb{R}$ is the corresponding label.

1. (**Data Fitting**, `Short, 5 points`) We want to construct a linear predictive model $\mathbf{a} \cdot \mathbf{x}$ to fit the value y. One possible loss function of this fitting is the worst-case error, i.e., $\max_i |\mathbf{a} \cdot \mathbf{x}_i - y_i|$. Show that computing the linear predictive model that minimizes the worst-case error can be formulated as a linear program.

2. (**Linear Classification**, `Short, 5 points`) When $y_i \in \{-1, 1\}$ is a binary label for data point i, this gives rise to a binary classification problem. In linear classification, we seek to find a hyperplane $\mathbf{a} \cdot \mathbf{x} - b = 0$ that strictly separates the data points with label 1 from the points with label -1. That is, $\mathbf{a} \cdot \mathbf{x}_i - b > 0$ if $y_i = 1$ and $\mathbf{a} \cdot \mathbf{x}_i - b < 0$ if $y_i = -1$ (note that the requirement "$<$" or "$>$" is strict). For convenience, we also call such a hyperplane *separating hyperplane*. Show that computing a separating hyperplane or asserting that it does not exist can be formulated as a linear feasibility problem (i.e,, a linear program with an arbitrary objective function).

   (Note that in linear programs, any linear constraint *cannot* have strict inequalities like "$>$" or "$<$"; see the general form of LPs in Lecture 2 slides.)

Additional Problems for those who take the course as Elective

Notes for regular students: *if you are interested*, you can try to solve these problems as well, but this is not a requirement for you. We will also grade your solution just *by courtesy*, but your grades on these problems will NOT count towards your final grade.

Problem 4: Optimal Strategic Attack to ML Algorithms (`Long, 15 points`)

Strategic or adversarial attacks to machine learning algorithms has been a hot research topic recently. In this question, you will device an *optimal strategic attack* to the machine learning algorithm discussed in Lecture 1 of our class. In particular, we studied the problem of selling a product (with unlimited supply) to N sequentially arriving buyers. All the buyers have the same value $v \in [0, 1]$ for the product but v is unknown to the seller. In order to maximize the seller's revenue, we described an online learning algorithm for selling the product that achieves regret $(2 \log \log N + 1)$.

In this question, we concern a slight variant of the above problem. That is, the seller sells the product (with unlimited supply) to a *single buyer* who repeatedly shows up for N rounds. The buyer's value $v \in [0, 1]$ (assume $v > 1/N$) is unknown to the seller. We assume that the seller still uses exactly the same algorithm as we described in class (you may need to review lecture 1 slides if necessary). Naturally, knowing that the seller is learning his value, the buyer will be strategic about his response at each round. For example, when offered a price p_n at round n, the buyer may intentionally respond with "Reject" even though $v > p_n$ because this will trick the seller to offer lower prices in next rounds. On the other hand, a "Reject" response also leads to buyer utility 0 whereas an "Accept" could have given him a utility of $v - p_n (> 0)$ at round n. Therefore, the strategic buyer who looks to *maximize his total utility* would need to balance between using "Reject" to induce lower prices and using "Accept" to collect positive utilities.

More formally, denote the seller's price at round n by $p_n \in [0, 1]$ and the buyer's response by $s_n \in \{0, 1\}$ for $n = 1, \cdots, N$, where $s_n = 1$ means the buyer responds with "Accept" at round n and $s_n = 0$ means a buyer response of "Reject". The total utility of the buyer is thus $\sum_{n=1}^{N} s_n(v - p_n)$. Suppose $v > 1/N$ and N is large enough. Assume that the seller is committed to run the algorithm as described in class (the one achieving $(2 \log \log N + 1)$ regret for N repeated buyers), what is the optimal buyer response strategy $s = (s_1, \cdots, s_N)$? Please describe an $O(N)$ time algorithm to compute the optimal buyer strategy.

What if the seller runs the standard binary search algorithm? How to compute the optimal buyer response strategy? Will the buyer gain less or more utility in this case? Prove your answers.

Hint: think about the following question — if the buyer will reject the offer for k rounds for some $k \leq N$, which k of the N rounds should he choose to reject the offer so that it maximizes his utility?