Note: Your TA may not get to all the problems. This is totally fine, the discussion worksheets are not designed to be finished in an hour. The discussion worksheet is also a resource you can use to practice, reinforce, and build upon concepts discussed in lecture, readings, and the homework.

Zero Sum Games: In this game, there are two players: a maximizer and a minimizer. We generally write the payoff matrix M in perspective of the maximizer, so every row corresponds to an action that the maximizer can take, every column corresponds to an action that the minimizer can take, and a positive entry corresponds to the maximizer winning. M is a n by m matrix, where n is the number of choices the maximizer has, and m is the number of choices the minimizer has.

A linear program that represents fixing the maximizer's choices to a probabilistic distribution where the maximizer has n choices, and the probability that the maximizer chooses choice i is p_i is the following:

$$\begin{gathered}
\max(z) \\
M_{1,1}(p_1) + \cdots + M_{n,1}(p_n) \geq z \\
M_{1,2}(p_1) + \cdots + M_{n,2}(p_n) \geq z \\
\vdots \\
M_{1,n}(p_1) + \cdots + M_{n,n}(p_n) \geq z \\
p_1 + p_2 + \cdots + p_n = 1 \\
p_1, p_2, \cdots, p_n \geq 0
\end{gathered}$$

The dual represents fixing the minimizers choices to a probabilistic distribution.

By strong duality, the optimal value of the game is the same if you fix the minimizer's distribution first or the maximizer's distribution first.

1 Weighted Rock-Paper-Scissors

You and your friend used to play rock-paper-scissors, and have the loser pay the winner 1 dollar. However, you then learned in CS170 that the best strategy is to pick each move uniformly at random, which took all the fun out of the game.
Your friend, trying to make the game interesting again, suggests playing the following variant: If you win by beating rock with paper, you get 5 dollars from your opponent. If you win by beating scissors with rock, you get 3 dollars. If you win by beating paper with scissors, you get 1 dollar.

(a) Draw the payoff matrix for this game.

(b) Write a linear program to find the optimal strategy.

Solution:

(a)

		Your Friend:		
		rock	paper	scissors
You:	rock	0	-5	3
	paper	5	0	-1
	scissors	-3	1	0

(b) Let r, p, s be the probabilities that you play rock, paper, scissors respectively. Let z stand for the expected payoff, if your opponent plays optimally as well.

$$\begin{aligned}
\max \quad & z & \\
5p - 3s &\geq z & \text{(Opponent chooses rock)} \\
s - 5r &\geq z & \text{(Opponent chooses paper)} \\
3r - p &\geq z & \text{(Opponent chooses scissors)} \\
r + p + s &= 1 & \\
r, p, s &\geq 0 &
\end{aligned}$$

2 Domination

In this problem, we explore a concept called *dominated strategies*. Consider a zero-sum game with the following payoff matrix for the row player:

Column:

		A	B	C
	D	1	2	-3
Row:	E	3	2	-2
	F	-1	-2	2

(a) If the row player plays optimally, can you find the probability that they pick D without directly solving for the optimal strategy? (Hint: Notice that the payoff for E is always greater than the payoff for D. When this happens, we say that E *dominates* D, i.e. D is a *dominated strategy*).

(b) Given the answer to part a, if the both players play optimally, what is the probability that the column player picks A?

(c) Given the answers to part a and b, what are both players' optimal strategies? (You might be able to figure this out without writing or solving any LP).

Solution:

(a) 0. Regardless of what option the column player chooses, the row player always gets a higher payoff picking E than D, so any strategy that involves a non-zero probability of picking D can be improved by instead picking E.

(b) 0. We know that the row player is never going to pick D, i.e. will always pick either E or F. But in this case, picking B is always better for the column player than picking A (A is only better if the row player picks D). That is, conditioned on the row player playing optimally, B dominates A.

(c) Based on the previous two parts, we only have to consider the probabilities the row player picks E or F and the column player picks B or C. Looking at the 2-by-2 submatrix corresponding to these options, it follows that the optimal strategy for the row player is to pick E and F with probability 1/2, and similarly the column player should pick B, C with probability 1/2.

3 Multiplicative Weights Intro

Multiplicative Weights
This is an online algorithm, in which you take into account the advice of n experts. Every day you get more information on how good every expert is until the last day T.
Let's first define some terminology:

- $x_i^{(t)}$ = proportion that you 'trust' expert i on day t
- $l_i^{(t)}$ = loss you would incur on day i if you invested everything into expert i
- total regret: $R_T = \sum_{t=1}^{T} \sum_{i=1}^{n} x_i^{(t)} l_i^{(t)} - \min_{i=1,\ldots,n} \sum_{t=1}^{T} l_i^{(t)}$

$\forall i \in [1, n]$ and $\forall t \in [1, T]$, the multiplicative update is as follows:

$$w_i^{(0)} = 1$$

$$w_i^{(t)} = w_i^{(t-1)} (1 - \epsilon)^{l_i^{(t-1)}}$$

$$x_i^{(t)} = \frac{w_i^{(t)}}{\sum_{i=1}^{n} w_i^{(t)}}$$

If $\epsilon \in (0, 1/2]$, and $l_i^{(t)} \in [0, 1]$, we get the following bound on total regret:

$$R_T \leq \epsilon T + \frac{\ln(n)}{\epsilon}$$

Let's play around with some of these questions. For this problem, we will be running the randomized multiplicative weights algorithm with two experts. Consider every subpart of this problem distinct from the others.

(a) Let's say we believe the best expert will have cost 20, we run the algorithm for 100 days, and epsilon is $\frac{1}{2}$. What is the maximum value that the total loss incurred by the algorithm can be?

(b) What value of ϵ should we choose to minimize the total regret, given that we run the algorithm for 25 days?

(c) We run the randomized multiplicative weights algorithm with two experts. In all of the first 140 days, Expert 1 has cost 0 and Expert 2 has cost 1. If we chose $\epsilon = 0.01$, on the 141st day with what probability will we play Expert 1? (Hint: You can assume that $0.99^{70} = \frac{1}{2}$)

Solution:

(a) total regret = loss of algorithm - offline optimum $\leq \epsilon T + \frac{\ln(n)}{\epsilon}$.

$$\text{loss of algorithm} - 20 \leq \frac{1}{2}(100) + \frac{\ln(2)}{\frac{1}{2}}$$

$$\text{loss of algorithm} \leq 50 + 2\ln(2) + 20$$

The maximum loss is roughly 71.39.

(b)

$$R_T \le \epsilon T + \frac{\ln(n)}{\epsilon}$$

$$R_T \le \epsilon 25 + \frac{\ln(2)}{\epsilon}$$

Take the derivative with respect to epsilon and set the derivative to 0 to get:

$$25 - \frac{\ln(2)}{\epsilon^2} = 0$$

$$25 = \frac{\ln(2)}{\epsilon^2}$$

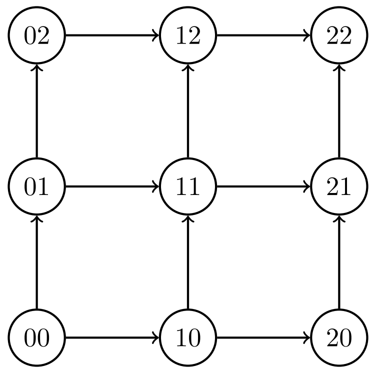
$$\epsilon = \sqrt{\frac{\ln(2)}{25}}$$

ϵ should be roughly 0.17.

(c) The weight assigned to expert 1 is $.99^{0 \cdot 140} = 1$, while the weight assigned to expert 2 is $.99^{1 \cdot 140} \approx 1/4$. So, the probability we play expert 1 is $\frac{1}{1+1/4} = 4/5$.

4 Multiplicative Weights

Consider the following simplified map of Berkeley. Due to traffic, the time it takes to traverse a given path can change each day. Specifically, the length of each edge in the network is a number between $[0, 1]$ that changes each day. The travel time for a path on a given day is the sum of the edges along the path.



For T days, both Max and Vinay drive from node 00 to node 22.
To cope with the unpredictability of traffic, Vinay builds a time machine and travels forward in time to determine the traffic on each edge on every day. Using this information, Vinay picks the path that has the smallest total travel time over T days, and uses the same path each day.
Max wants to use the multiplicative weights update algorithm to pick a path each day. In particular, Max wants to ensure that the difference between his expected total travel time over T days and Vinay's total travel time is at most $T/10000$. Assume that Max finds out the lengths of all the edges in the network, even those he did not drive on, at the end of each day.

(a) How many experts should Max use in the multiplicative weights algorithm?

(b) What are the experts?

(c) Given the weights maintained by the algorithm, how does Max pick a route on any given day?

(d) The regret bound for multiplicative weights is as follows:

Theorem. Assuming that all losses for the n experts are in the range $[0, 4]$, the worst possible regret of the multiplicative weights algorithm run for T steps is

$$R_T \leq 8\sqrt{T \ln n}$$

Use the regret bound to show that expected total travel time of Max is not more than $T/10000$ worse than that of Vinay for large enough T.

Solution:

(a) 6 experts

(b) There is one expert for every path from 00 to 22. One can see that there are 6 different paths from 00 to 22.

(c) The algorithm maintains one weight for each path. Max picks a path with probability proportional to its weight.

(d) Let $P_1, \ldots, P_6$ be the 6 possible paths between 00 and 22. Let $\ell_i^{(t)}$ denote the length of path i on day t. Let $w_i^{(t)}$ denote the weight of path i on day t.

Since Max picks a path proportional to its weight, the expected total time on day t is

$$\sum_{i=1}^{6} w_i^{(t)} \cdot \ell_i^{(t)},$$

and the expected total time over T days is,

$$\sum_{t=1}^{T} \sum_{i=1}^{6} w_i^{(t)} \cdot \ell_i^{(t)}.$$

The regret bound to multiplicative weights asserts that for every path P_i,

$$\sum_{t=1}^{T} \sum_{i=1}^{6} w_i^{(t)} \cdot \ell_i^{(t)} - \sum_{t=1}^{T} \ell_i^{(t)} \leq 8\sqrt{T \ln 6}\ .$$

Here $n = 6$, since there are 6 different paths. Since Vinay picks one of the paths P_i to use over all days, the above regret bound implies that total expected time of Max is at most $8\sqrt{T \ln 6}$ worse than that of Vinay. For sufficiently large T, $8\sqrt{T \ln 6} < T/10000$, in particular $T > (8 \cdot 10000)^2 \cdot \ln 6$ suffices.