INSTRUCTIONS

- **Due: Friday, 6 December 2019 at 10:00 PM EDT.** Remember that you have NO slip days for Written Homework, but you may turn it in up to 24 hours late with 50% Penalty.
- **Format:** Submit the answer sheet pdf containing your answers. You should solve the questions on this handout (either through a pdf annotator, or by printing, then scanning). Make sure that your answers (typed or handwritten) are within the dedicated regions for each question/part. If you do not follow this format, we may deduct points.
- **How to submit:** Submit a pdf with your answers on Gradescope. Log in and click on our class 15-281 and click on the submission titled HW12 and upload your pdf containing your answers.
- **Policy:** See the course website for homework policies and Academic Integrity.

Name	
Andrew ID	
Hours to complete?	

For staff use only

Q1	Q2	Q3	Q4	Total
/20	/25	/20	/35	/100

Q1. [20 pts] "Pick Your Favorite" Rock Paper Scissors

Suppose Michelle and Vicky are playing rock paper scissors (RPS). Besides the winning and losing schemes of traditional RPS, both Michelle and Vicky have their own favorite choice among the three available choices that makes them happier when they win and more discouraged when they lose. The payoff matrix for this game is shown below.

M \ V	R	P	S
R	(2, 0)	(-4, 8)	(6, -1)
P	(1, -1)	(0, 2)	(-1, 1)
S	(-1, 1)	(1, -6)	(0, 0)

In each grid of the table, the first number denotes Michelle's payoff, and the second number denotes Vicky's payoff.

(a) [5 pts] (i) In the game described above, is it possible to have a Nash Equilibrium in the form of $(0,\, p,\, 1-p)$ and $(q,\, 0,\, 1-q)$ as Michelle's and Vicky's mixed strategies respectively (where $0 < p < 1$ and $0 < q < 1$, p is the probability of Michelle playing Paper and q is the probability of Vicky playing Rock)?

○ Yes ○ No

(ii) Explain why or why not.

Answer:

(b) [15 pts] You were told that the game has a Nash Equilibrium in the form of $(p_1,\, p_2,\, 1-p_1-p_2)$ and $(q_1,\, q_2,\, 1-q_1-q_2)$ respectively, where $0 < p_1,\, p_2 < 1$ and $0 < q_1,\, q_2 < 1$ (p_1 is the probability of Michelle playing Rock and q_1 is the probability of Vicky playing Rock). What are Michelle and Vicky's strategies in this equilibrium? Show all steps of your work.

Steps:

Q2. [25 pts] Linear Programming for Game

Consider the following table representing a zero-sum game.

P2 P1	A1	A2	A3
A1	(-4, 4)	(0, 0)	(3, -3)
A2	(6, -6)	(1, -1)	(-5, 5)

A1, A2 and A3 are three actions for both players. P1 is the row player and P2 is the column player. In each cell, the first number of the tuple represents the payoff of P1, and the second number of the tuple represents the payoff of P2.

We want to represent this game using linear programming with Maximin Strategy. Below is an incomplete LP with respect to P1's view:

$$\begin{aligned} max_{V,p_1,p_2} V \\ s.t. p_1 + p_2 \leq 1 \\ -4p_1 + 6p_2 \geq V \\ p_2 \geq V \\ 3p_1 - 5p_2 \geq V \end{aligned}$$

Here we use V to represent P1's value in the worst case. p_1 and p_2 are the probabilities of P1 choosing A1, A2 respectively. Similar logic goes for q_1, q_2 and q_3, which are probabilities for P2.

(a) [10 pts] Write the missing constraints in the formulation above.

Answer:

(b) [15 pts] Write a linear programming formulation with respect to P2.

Answer:

Q3. [20 pts] "Top Three" Voting Rule

Consider a new voting rule. Under this rule, every voter's top three preferences get one point each. The outcome with most points win. Assume ties are broken in alphabetical order, e.g. *b* is chosen over *c*.

Answer the following questions given this preference profile:

Category 1	Category 2	Category 3
10 voters	35 voters	45 voters
a	f	e
b	c	f
c	d	d
e	b	a
f	e	c
d	a	b

(a) [3 pts] Who is the winner?

Answer:

(b) [15 pts] Can a player in the first category manipulate the voting to make another player the unique winner (without tie-breaking)? If yes, what is his reported preference according to the greedy algorithm? If not, briefly explain why.

○ Yes ○ No

State preference / Explain why not:

(c) [12 pts] Which properties does this voting rule satisfy?

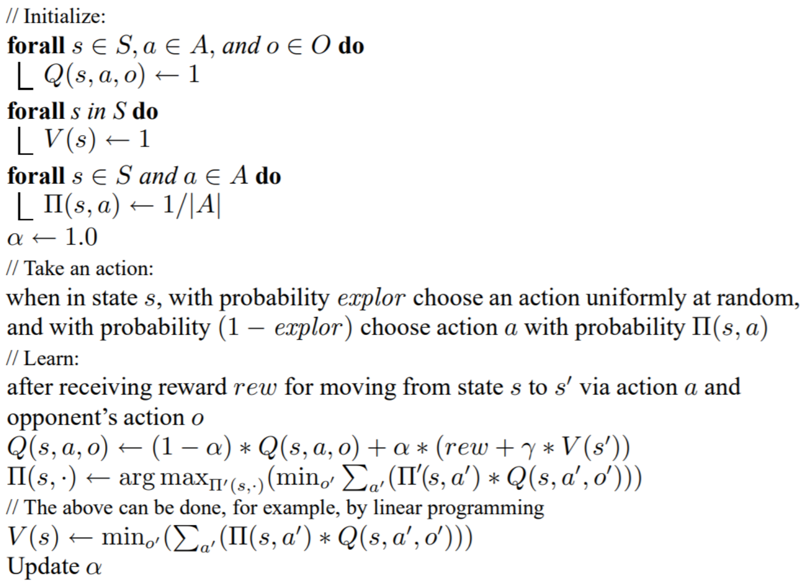
- □ Majority Consistency
- □ Condorcet Consistency
- □ Strategy-proof
- □ Dictatorial
- □ Constant
- □ Onto

Q4. [35 pts] Minimax-Q learning

Bert and Ernie from Recitation 11 are back for more entertainment. There are still two lanes, A and B. This time, however, they decide to play for multiple rounds. In each round, they both start in lane A, and they can choose to stay in lane A or switch to lane B. They'll crash if they meet while in the same lane, continuing the next round with backup cars. There is a change in reward. Ernie has learned that she won't be getting a new lambo unless her current one breaks. Therefore, Ernie actually wants to crash while Bert doesn't want to. The reward table is updated as follows:

Ernie Bert	A	B
A	-5,5	3,-3
B	3,-3	-5,5

You may want to read up Chapter 7.4 from Shoham & Leyton-Brown on minimax-Q learning (`http://www.masfoundations.org/mas.pdf`). For convenience, below is the extracted pseudocode:



```
// Initialize:
forall $s \in S, a \in A$, and $o \in O$ do
    $Q(s,a,o) \leftarrow 1$
forall $s$ in $S$ do
    $V(s) \leftarrow 1$
forall $s \in S$ and $a \in A$ do
    $\Pi(s,a) \leftarrow 1/|A|$
$\alpha \leftarrow 1.0$
// Take an action:
when in state $s$, with probability $explor$ choose an action uniformly at random,
and with probability $(1 - explor)$ choose action $a$ with probability $\Pi(s,a)$
// Learn:
after receiving reward $rew$ for moving from state $s$ to $s'$ via action $a$ and
opponent's action $o$
$Q(s,a,o) \leftarrow (1-\alpha) * Q(s,a,o) + \alpha * (rew + \gamma * V(s'))$
$\Pi(s,\cdot) \leftarrow \arg\max_{\Pi'(s,\cdot)}(\min_{o'} \sum_{a'}(\Pi'(s,a') * Q(s,a',o')))$
// The above can be done, for example, by linear programming
$V(s) \leftarrow \min_{o'}(\sum_{a'}(\Pi(s,a') * Q(s,a',o')))$
Update $\alpha$
```

Let $\gamma = 0.9$. You are given the α and chosen actions in each round. For simplicity, let the values for states A and B be equivalent (so $Q(A,i,j) = Q(B,i,j), V(A) = V(B)$). A mixed strategy π(A) can be represented as $(x, 1-x)$, where x = P(stay in lane A), while $1 - x$ = P(switch to lane B). Bert and Ernie can have different strategies.
Fill in the **bolded blue** values for Bert in the table below, and show your work on the next page. Note that some entries may have multiple correct answers, providing one is sufficient.

Round	α	Actions	Reward	Q(A,st,st)	Q(A,st,sw)	Q(A,sw,st)	Q(A,sw,sw)	$\pi(A)$	V(A)
0	1			1	1	1	1	(0.5,0.5)	1
1	1	(sw, st)	3	1	1	**H**	1	**I**	**J**
2	1	(st, st)	-5	**K**	1	H	1	**L**	1
3	1	(st, sw)	3	K	**M**	H	1	(0.266, 0.734)	**N**

Notation:
Actions (sw, st): Bert's action = switch, Ernie's action = stay
Q(A, st, sw): Q-value at state = lane A, Bert's action = stay, Ernie's action = switch.

1. **H: Q(A, sw, st) in Round 1:**

2. **I: π(A) in Round 1:**

3. **J: V(A) in Round 1:**

4. **K: Q(A, st, st) in Round 2:**

5. **L: π(A) in Round 2:**

6. **M: Q(A, st, sw) in Round 3:**

7. **N: V(A) in Round 3:**