MAS Final Exam.
Solution Sheet

Question 1.

10/10

	L	C	R
U	2 0	1 1	4, 2
M	3 4	1 2	2 3
D	1 3	0 2	3 0

① Strategies that survive IESDS.

	2 0	1 1	4 2
	3 4	1 2	2 3

	L	R
U	2 0	4 2
M	3 4	2 3

2/2

(U, M), (L, R)

player 1   player 2

1.2 Pure Strategy NE.

	L	R
U	2 0	4 2
M	3 4	2 3

PNE: (M, L) → 3, 4

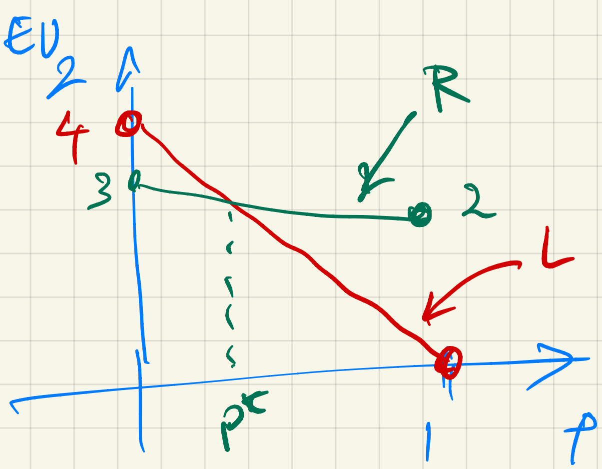
(U, R) → 4, 2.

2/2

1.3: Mixed NE

		L(q)	R(1-q)
p	U	2 0	4, 2
1-p	M	3, 4	2, 3

3/3



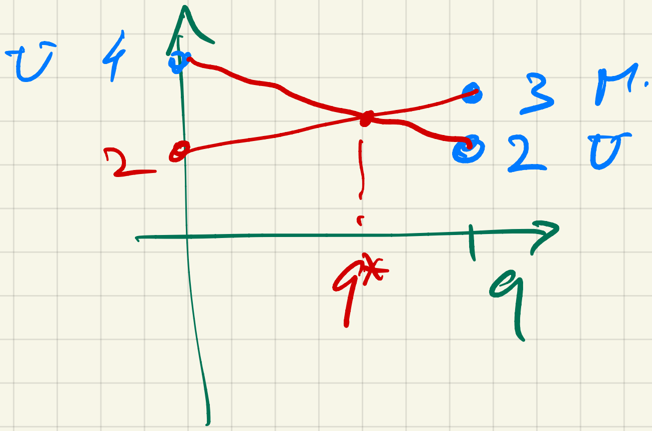
$$EU_2(p, L) = EU_2(p, R)$$

$$0 \cdot p + 4(1-p) = 2p + 3(1-p)$$

$$4 - 4p = 2p + 3 - 3p$$

$$3p = 1 \Rightarrow p^* = 1/3$$

$q^* = ?$



$$EU_1(U,q) = EU_1(M,q)$$

$$2q + 4(1-q) = 3q + 2(1-q)$$

$$2q + 4 - 4q = 3q + 2 - 2q = q + 2$$

$$-2q + 4 = q + 2$$

$$3q = 2 \Rightarrow \boxed{q^* = 2/3}$$

---

4. Expected utility

MNE: $p^* \to U$, $1-p^* \to M$

$q^* \to L$, $1-q^* \to R$.

$p^* = 1/3$, $q^* = 2/3$

2/2

	q^*	$1-q^*$
p^*	2 0	4 2
$1-p^*$	3 4	2 3

	$2/3$	$1/3$
$p^*=1/3$	$2/9$	$1/9$
$2/3$	$4/9$	$2/9$

$$EU_1(p^*,q^*) = \frac{1}{9}(2.2 + 1.4 + 3.4 + 2.2)$$

$$= \frac{1}{9}(24) = \frac{24}{9} = \frac{8}{3}$$

$$EU_2(p^*,q^*) = \frac{1}{9}(0.2 + 2.1 + 4.4 + 2.3)$$

$$= \frac{1}{9}(24) = \frac{8}{3}$$

$$\underline{\text{PNE}}: \left.\begin{array}{l} (M,L) \longrightarrow 3,4. \\ (U,R) \longrightarrow 4,2 \end{array}\right\} \text{utilities from table.}$$

<u>1.5</u>: Prediction impossible because there are three NE.

1/1

2. Game Theory: Investment game → 5/5

2.1 Majority situation → 2/2

Investors

Majority

Eg

20 0
x x x x o o o
invest | no invest

wants to change.
utility: 0 → 20.

x x x x x o o
still majority

Stable if all invest

NE
x x x x x x x
all invest.

Minority.

−10 0
x x x o o o o

wants to change
utility: −10 → 0

Only stable situation:

o o o o o o o

None invest

NE: no investors.

Conclusion: 2 NE → all, → none.

2.2 : Minority game. → 3/3

Investors get pos. reward when in minority

Investors

Majority

−10 | 0

x x x x x | o o

change

$-10 \to 0$

not stable.

x x x x o o o

change:

$-10 \to 0$

No NE in this case.

Minority

① 20 0

x x o o o o o

wants to change!

util: $0 \to 20$

② 20 0

x x x o o o o

change!

util: $20 \to 0$

change.

util: $0 \to -10$

None wants to change

NE

max minority.

util: $0 \to 20$.

Conclusion

in minority game

NE: max minority

i.e. $n = 2k+1 \longrightarrow$ max minority $= k$.

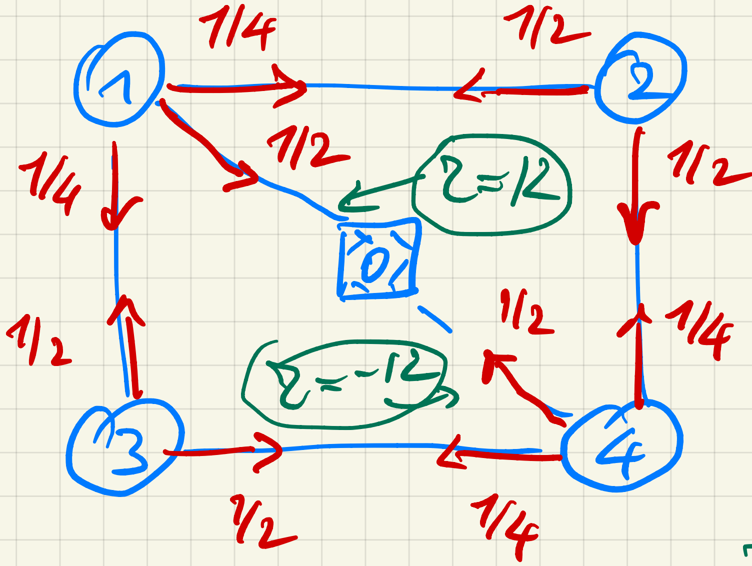
$$\underbrace{x \cdots\cdots x}_{k} \quad \underbrace{o \; o \cdots\cdots o}_{k+1}$$

invest   not invest.

3. MDP → 10/10

π : policy

①



TO ↙

$$
P_\pi = \begin{array}{c|ccccc}
 & 0 & 1 & 2 & 3 & 4 \\
\hline
0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1/2 & 0 & 1/4 & 1/4 & 0 \\
2 & 0 & 1/2 & 0 & 0 & 1/2 \\
3 & 0 & 1/2 & 0 & 0 & 1/2 \\
4 & 1/2 & 0 & 1/4 & 1/4 & 0
\end{array}
$$

FROM →

2/2

$$0, \quad 11/2, \quad -1, \quad -1, \quad -13/2$$

$$r_\pi = \left(0, \tfrac{1}{2}(12-1), -1, -1, \tfrac{1}{2}(-12-1)\right)$$

2/2

0 1 2 3 4

2) Optimal v^* $\quad \gamma = 2/3$ $\quad$ 2/2

optimal policy:

go to 0 via 1 asap.

Bellman optimality eq:

$$v^*(s) = \max_a \left(r(s,a,s') + \gamma v^*(s') \right)$$

$v^*(0) = 0.$

$$\begin{cases} v^*(1) = 12 \\ v^*(2) = -1 + \frac{2}{3} \cdot 12 = -1 + 8 = 7. \\ v^*(3) = 7. \\ v^*(4) = -1 + \frac{2}{3} \cdot 7 = \frac{11}{3} \end{cases}$$

Policy not unique since we can choose in 4.

1/1 ③ $q^*(1,a)$ when $1 \xrightarrow{a} 2$

(assume: $\gamma = 2/3$)

$$q^*(s,a) = r(s,a,s') + \gamma v^*(s')$$

$$= r(s,a,s') + \gamma \max_{a'} q(s',a')$$

$$q^*(1,a) = r(1,a,2) + \gamma v^*(2)$$

$$= -1 + \frac{2}{3}\, 7 = \frac{11}{3}$$

④ Non-deterministic transitions

3/3 $P_\pi(1,:) = \boxed{\begin{pmatrix} \frac{3}{8} & \frac{1}{4} & \frac{3}{16} & \frac{3}{16} & 0 \end{pmatrix}}$

$$\begin{array}{c|ccccc|} 1 & \frac{3}{4}\,\frac{1}{2} & 1/4 & \frac{3}{4}\cdot\frac{1}{4} & \frac{3}{4}\cdot\frac{1}{4} & 0 \\ \hline & 0 & 1 & 2 & 3 & 4 \end{array}$$

$$v_\pi(1) = \frac{1}{4}(-2) + \frac{3}{4}\left(\frac{1}{2}\cdot 12 - \frac{1}{2}1\right)$$

$$= -\frac{1}{2} + \frac{3}{4}\left(\frac{11}{2}\right)$$

$$= \frac{-4+33}{8} = \frac{29}{8}.$$

$$v_\pi(2) = \frac{1}{4}(-2) + \frac{3}{4}(-1) = -\frac{5}{4}.$$

4. RL & Exploration vs. Exploitation

4.1 Q-learning.

s = State	a	s'	r	q(s,a)	v(s)	π(a|s)
2	R	3	-1	8	5	1/4
2	L	1	0	4	5	3/4
3	R	4	1	6	7	2/3
3	L	2	-2	9	7	1/3

Computation Q-values:

$$\boxed{v(s) = \sum_a q(s,a)\,\pi(a|s)}$$

1. $5 = \frac{1}{4} q(2,R) + \frac{3}{4} \cdot 4 = \frac{1}{4} q(2,R) + 3$

$\boxed{q(2,R) = 8}$ $\quad 11/3 = 3.67.$ alternative (see next page)

2. $7 = 6 \cdot \frac{2}{3} + q(3,L) \cdot \frac{1}{3} \Rightarrow (\underbrace{7-4}_{3}) \cdot 3 = q(3,L)$

$\boxed{q(3,L) = 9}$ $\quad 1.33 = 4/3$

Alternative computation of $\begin{cases} q(2,R) \\ q(3,L) \end{cases}$

Deterministic transitions:

$$s \xrightarrow{a} s'$$

$$\boxed{q(s,a) = r(s,a,s') + \gamma v(s')}$$

Hence, $2 \xrightarrow{R} 3$

1. $q(2,R) = -1 + \frac{2}{3} v(3) = -1 + \frac{2}{3} \cdot 7 = \frac{11}{3}$

$$= 3.67.$$

$$3 \xrightarrow{L} 2$$

2. $q(3,L) = -2 + \frac{2}{3} v(2) = -2 + \frac{2}{3} 5$

$$= \frac{4}{3} = 1.33$$

$2 \to 3$

1 $q_\pi'(2,R) \leftarrow q_\pi(2,R) + \alpha \left[r + \gamma \max_{a'} q(3,a') - q(2,R) \right]$

$8 + 0.9 \left[-1 + \frac{2}{3} \underbrace{\max(6,9)}_{9} - 8 \right]$

$\underbrace{}_{9}$, $\underbrace{}_{6}$, $\underbrace{}_{5}$ $\to -3$

$= 8 - 3.(0.9) = 8 - 2.7 = 5.3$

3/3 = q-values + update

Alternative $= 3.67 + 0.9 \left[-1 + \underbrace{\frac{2}{3} \max(6, 1.33)}_{3} - 3.67 \right] = 2.7 + 0.37 = 3.07$

2 SARSA

2/2

$s \xrightarrow{a} s' \xrightarrow{a'} \dots$

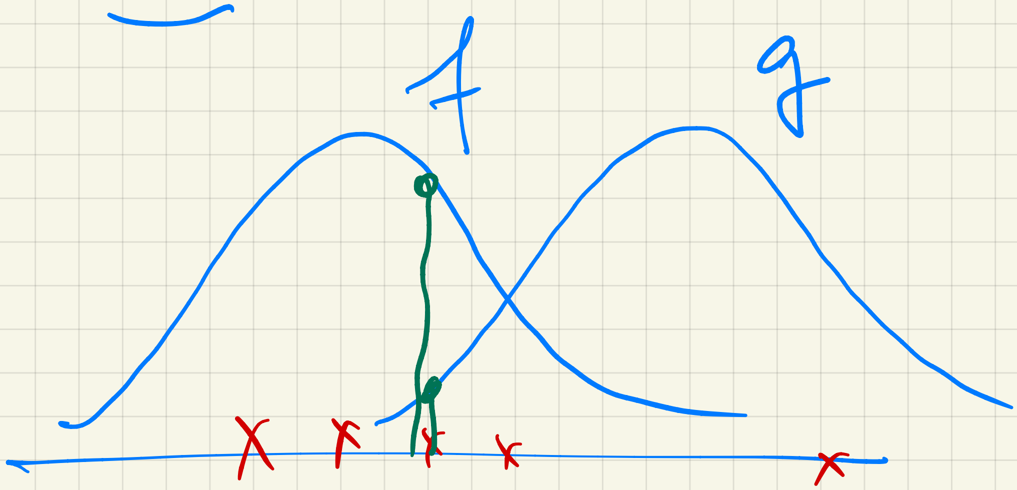
need to know a' !

3 See theory.

2/2

4.2 KL.

3 | 3



$X_i \sim f$ $\rightarrow$ more sample points near high f-values

$\downarrow$

for those, $\dfrac{f(X_i)}{g(X_i)} > 1$

$$\Rightarrow \log \frac{f(X_i)}{g(X_i)} > 0$$