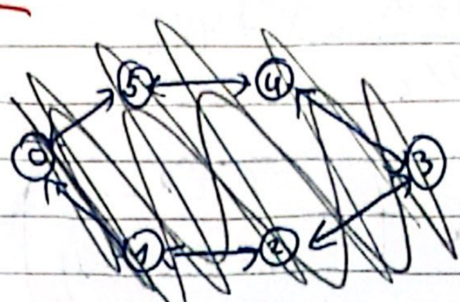
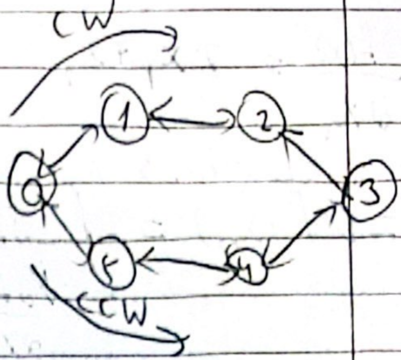
RL - EX2

Question 1:

1.





$S = \{0, 1, \dots, 2k-1\}$ - State space

$A = \{CW, CCW\} = \{i \text{ to } (i+1)\%k, \ i \text{ to } (i-1)\%k\}$ - Action state

P = probability matrix that action a in state s at time t will lead to state s' at t+1. $Pr(S_{t+1} = s' | S_t = s, a_t =$

$$Pr(S_{t+1} = s' | S_t = s, a_t = CW) = \begin{cases} 1 & s' = (s+1)\%2k \\ 0 & o.w. \end{cases}$$

$$Pr(S_{t+1} = s' | S_t = s, a_t = CCW) = \begin{cases} 1 & s' = (s-1)\%2k \\ 0 & o.w. \end{cases}$$

R = expected reward received for transitioning from s to s', due to a
$R_a(s, s')$.

$$= \begin{cases} 1 & s' = 0, \forall s \in \{1, 2k-1\} \ \forall a \in \{cw, ccw\} \\ 0 & o.w \end{cases}$$

$S_0 = k$ (by the definition of the problem)

2. $$\pi^* = \begin{cases} a_t = ccw \text{ if } s_t \in \{0,1,\dots,k\} \\ a_t = cw \text{ if } s_t \in \{k+1,\dots,2k-1\} \end{cases}$$

המדיניות האופטימלית היא כזו שתביא אותנו ל-0 כמה שיותר ~~מהר~~ מהר,
וגם כשהגענו ל-0 ונאלץ לעזוב אותו, נחזור אליו מיד עם הפעולה ההפוכה.
במקרה שלנו, מתחילים מ $S_0=k$ עד $S_{t=k}=0$ כי הולכים נגד
השעון, עוד צעד נוסף נגד השעון יוביל ל $S_{t=k+1}=2k-1$ ולכן נהפוך
את הפעולה ונלך עם השעון ל $S_{t=k+2}=0$ וחוזר חלילה.

3. $$V_{k+1}(s) = \max_{a \in \{cw,ccw\}} \left[r_k(s,a) + \gamma \sum_{s' \in S_{k+1}} p(s'|s,a) V_k(s') \right]$$

$$V_1(s) = \max_a r(s,a) \quad \text{אז} \quad V_0(s) = 0 \ \forall s$$

$$V_1(s=0) = 1, \ V_1(s \neq 0) = 0$$

רק $V_1(0)$ השתנה (עבור מצב 0)

4. $$V_2(s) = \max_a \left[r(s,a) + \gamma \sum_{s' \in S_2} p(s'|s,a) V_1(s') \right]$$

if $s'=0$: ($s'=0 \in S_2$ only for $s \in \{1, 2k-1\}$)

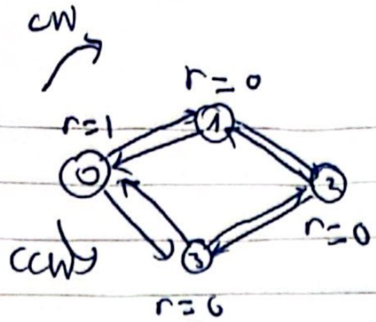
$$V_2(s \in \{1,2k-1\}) = \max_a \left[r(s,a) + \gamma V_1(0) \right] = \gamma$$

if $s' \neq 0$ ($s' \neq 0 \in S_2$ is for $s \in S / \{1, 2k-1\}$)

$$V_2(s) = \max_a \left[r(s,a) + 0 \right] = \begin{cases} 1 & s=0 \\ 0 & s \neq 0,1,2k-1 \end{cases}$$

V-function השתנה עבור 2 מצבים נוספים: 1, 2k-1

5.



$V_1^*(s=0)=1 \quad , V_1^*(s=1,2,3)=0$ איטרציה 1:

$V_2^*(s=0)=1 \quad , V_2^*(s=\{1,3\})=\gamma \quad , V_2^*(s=2)=0$ איטרציה 2:

for s=1: $\max\big(\underbrace{0+\gamma\cdot V_1(s=0)}_{ccw},\ \underbrace{0+\gamma\cdot V_1(s=2)}_{cw}\big)=\gamma$

for s=3: $\cancel{\max\big(0+\gamma V_1(s=2),\ 0+\gamma V_1(s=0)\big)=\gamma}$

$\max\big(\underbrace{0+\gamma V_1(s=2)}_{ccw},\ \underbrace{0+\gamma V_1(s=0)}_{cw}\big)=\gamma$

$V_3^*(s=0)=\max\big(r(0,cw)+\gamma V_2(s=1),$ איטרציה 3:

$r(0,ccw)+\gamma V_2(s=3)\big)=1+\gamma^2$

$V_3^*(s=1)=\max\big(\underbrace{\gamma V_2(s=0)}_{ccw},\ \underbrace{\gamma V_2(s=2)}_{cw}\big)=\gamma$

$V_3^*(s=3)=\max\big(\underbrace{\gamma V_2(s=2)}_{ccw},\ \underbrace{\gamma V_2(s=0)}_{cw}\big)=\gamma$

$V_3^*(s=2)=\max\big(\gamma V_2(s=1),\ \gamma V_2(s=3)\big)=\gamma^2$

$V_4(0)=1+\gamma^2$ איטרציה 4:

$V_4(1)=\gamma(1+\gamma^2)=\gamma+\gamma^3$

$V_4(2)=\gamma^2$

$V_5(0) = 1 + \gamma(\gamma+\gamma^3) = 1+\gamma^2+\gamma^4$ איטרציה 5:

$V_5(1) = \gamma(1+\gamma^2) = \gamma+\gamma^3$

$V_5(2) = \gamma^2+\gamma^4$

$V_6(0) = 1+\gamma^2+\gamma^4$ איטרציה 6:

$V_6(1) = \gamma+\gamma^5+\gamma^3$

$V_6(2) = \gamma^2+\gamma^4$

$\vdots$

ceil($\frac{k}{2}$)

$$\boxed{\begin{aligned} V_k(0) &= \sum_{n=0}^{\lceil \frac{k}{2} \rceil -1} \gamma^{2n} \quad k>0, \text{ o.w. } 0 \\ V_k(1) = V_k(3) &= \sum_{n=0}^{\lfloor \frac{k}{2} \rfloor -1} \gamma^{2n+1} \quad k>1, \text{ o.w. } 0 \\ V(2) &= -1 + \sum_{n=0}^{\lceil \frac{k}{2} \rceil -1} \gamma^{2n} \quad k>0, \text{ o.w } 0 \end{aligned}}$$