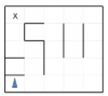
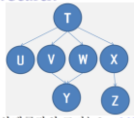
Section 2: Graph Search **Section 8**





1.U ㅗ X
2.U ㅗ X | T O
3.V ㅗ W | Y X
4.V ㅗ W | T O
5.T ㅗ Y | V X
6.Y ㅗ Z | W O
7.Y ㅗ Z | T O

독립X

MxN 공간이라면, 상태공간의 크기는? $4MN(V_{max}+1)$

맨하탄 거리 (agent → exit)가 admissible 한가? NO

Admissible heuristic을 정의해 보아라.

Relaxation (no walls, can turn and change speed) d맨하탄/Vmax

Section 5

1. Transition Function과 Reward Function은?

$T(s, Stop, Done) = 1$

$T(0, Draw, s') = 1/3$ for $s' \in \{2,3,4\}$

$T(2, Draw, s') = 1/3$ for $s' \in \{4,5, Done\}$

$T(3, Draw, s') = \begin{cases} 1/3 \text{ if } s' = 5 \\ 2/3 \text{ if } s' = Done \end{cases}$

$T(4, Draw, Done) = 1$

$R(s, Stop, Done) = s, s \le 5$

$R(s,a,s') = 0$ otherwise

$$V_{k+1}(s) \leftarrow \max_a \sum_{s'} T(s,a,s')\left[R(s,a,s') + \gamma V_k(s')\right]$$

States	0	2	3	4	5
V_0	0	0	0	0	0
V_1	0	2	3	4	5
V_2	3	3	3	4	5
V_3	10/3	3	3	4	5
V_4	10/3	3	3	4	5

States	0	2	3	4	5
π^*	Draw	Draw	Stop	Stop	Stop

$$\pi_{i+1}(s) = \arg\max_a \sum_{s'} T(s,a,s')\left[R(s,a,s') + \gamma V^{\pi_i}(s')\right]$$

States	0	2	3	4	5
π_i	Draw	Stop	Draw	Stop	Draw
V^{π_i}	2	2	0	4	0
π_{i+1}	Draw	Stop	Stop	Stop	Stop

$sample = R(s, \pi(s), s') + \gamma V^{\pi}(s')$

$V^{\pi}(s) \leftarrow V^{\pi}(s) + \alpha(sample - V^{\pi}(s))$

	A	B	C	D	E
0	0	0	0	0	0
1	0	1	0	0	0
2	0	1	2	0	0
3	0	1	4	0	0
4	0	3.5	4	0	0

(B, East, C, 2)
(C, South, E, 4)
(C, East, A, 6)
(B, East, C, 2)

Q-value from Q-Learning

$Q(B, East) = 3$
$Q(C, South) = 2$
$Q(C, East) = 3$
All other q-states have a value of 0.

Sec11_q3_4
Classify O
Classify X
Classify X

Section 6

$f_g = 2, f_p = 1$

$w_g = 100, w_p = -10,$

$Q(s,a) = w_g * f_g + w_p * f_p = 100*2 + -10*1 = 190$

$F_g=0, F_p=2,$ reward = 50

discount factor γ가 0.5

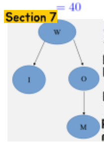
$$Q_{new}(s,a) = R(s,a,s') + \gamma * \max_{a'} Q(s',a')$$
$= 50 + 0.5*(100*0 + -10*2)$
$= 40$

예측 값과 학습률 (learning rate, α) 0.5를 이용

$w_g = w_g + \alpha * (Q_{new}(s,a) - Q(s,a)) * f_g(s,a)$
$= 100 + 0.5*(40-190)*2 = -50$

$w_p = w_p + \alpha * (Q_{new}(s,a) - Q(s,a)) * f_p(s,a)$
$= -10 + 0.5*(40-190)*1 = -85$

Section 7

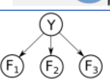


- 답 : Yes,
- I,W 사이의 화살표가 바뀐 network와 동치

Ex) Chain rule for probability
P(W, I, O, M) = P(W)P(I|W)P(O|W, I)P(M|W, I, O)

P(I, W, O, M) = P(I)P(W|I)P(O|W, I)P(M|W, I, O)

P(W)P(I|W) = P(I)P(W|I)이므로 I와 W의 화살표가 바뀐 network와 동치



F_1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0
F_2	0	0	1	0	0	0	0	1	0	0	1	0	0	1	1
F_3	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0
Y	A	B	B	B	B	B	B	B	B	B	B	B	B	C	C

$P(Y=A) = \frac{1}{15} = 0.067$
$P(Y=B) = \frac{12}{15} = 0.800$
$P(Y=C) = \frac{2}{15} = 0.133$

$P(F_1=0|Y=A) = \frac{1}{1} = 1.000$
$P(F_1=1|Y=A) = \frac{0}{1} = 0.000$
$P(F_1=0|Y=B) = \frac{0}{12} = 0.000$
$P(F_1=1|Y=B) = \frac{12}{12} = 1.000$
$P(F_1=0|Y=C) = \frac{2}{2} = 1.000$
$P(F_1=1|Y=C) = \frac{0}{2} = 0.000$

$P(Y=A, F_1=1, F_2=1, F_3=0) = 0.000*0.000*0.000*0.067 = 0.000$
$P(Y=B, F_1=1, F_2=1, F_3=0) = 1.000*0.250*1.000*0.800 = 0.200$
$P(Y=C, F_1=1, F_2=1, F_3=0) = 0.000*1.000*0.000*0.133 = 0.000$

$P(Y=A|F_1=1,F_2=1,F_3=0) = \frac{0.000}{0.000+0.200+0.000} = 0.000$
$P(Y=B|F_1=1,F_2=1,F_3=0) = \frac{0.200}{0.000+0.200+0.000} = 1.000$
$P(Y=C|F_1=1,F_2=1,F_3=0) = \frac{0.000}{0.000+0.200+0.000} = 0.000$

$P(F_1=0|Y=A) = \frac{1+4}{1+8} = 0.556$
$P(F_1=1|Y=A) = \frac{0+4}{1+8} = 0.444$
$P(F_1=0|Y=B) = \frac{0+4}{12+8} = 0.200$
$P(F_1=0|Y=B) = \frac{12+4}{12+8} = 0.800$
$P(F_1=0|Y=C) = \frac{2+4}{2+8} = 0.600$
$P(F_1=0|Y=C) = \frac{0+4}{2+8} = 0.400$

$P(Y=A) = \frac{1+4}{15+12} = 0.185$
$P(Y=B) = \frac{12+4}{15+12} = 0.593$
$P(Y=C) = \frac{2+4}{15+12} = 0.222$

$P(Y=A, F_1=1, F_2=1, F_3=0) = 0.444*0.444*0.444*0.185 = 0.016$
$P(Y=B, F_1=1, F_2=1, F_3=0) = 0.800*0.350*0.800*0.593 = 0.133$
$P(Y=C, F_1=1, F_2=1, F_3=0) = 0.400*0.600*0.400*0.222 = 0.021$

$P(Y=A|F_1=1,F_2=1,F_3=0) = \frac{0.016}{0.016+0.133+0.021} = 0.095$
$P(Y=B|F_1=1,F_2=1,F_3=0) = \frac{0.133}{0.016+0.133+0.021} = 0.779$
$P(Y=C|F_1=1,F_2=1,F_3=0) = \frac{0.021}{0.016+0.133+0.021} = 0.125$

P(T), P(U|T), P(V|T), P(W|T), P(X|T), P(Y|V,W), P(+z|X)

$f_1(+z,T) = \sum_x P(x|T)P(+z|x)$
$f_2(U,V,W,+z) = \sum_t P(t)P(U|t)P(V|t)P(W|t)f_1(+z,t)$
$f_3(V,W,+z) = \sum_u f_2(u,V,W,+z)$
$f_4(W,Y,+z) = \sum_v f_3(v,W,+z)P(Y|v,W)$
$f_5(Y,+z) = \sum_w f_4(w,Y,+z)$
$P(Y|+z) = \frac{f_5(Y,+z)}{z} = \frac{f_5(Y,+z)}{\sum_y f_5(y,+z)}$

큰 factor 사이즈
$f_2(U,V,W,+z)$ 8
Factor의 사이즈가 가장 작게 생성되는 방법
X, U, T, V, W

Section 9 (R,F,R,F), (R,F,R,F), (S,F,S,T), (S,T,S,T), (S,T,R,F), (R,F,R,T), (S,T,S,T), (S,T,S,T), (S,T,R,F), (R,F,S,T)

Likelihood weighting

$w = P(I_1=T|W_1=R) * P(I_2=F|W_2=R) = 0.9 * 0.8 = 0.72$
$w = P(I_1=T|W_1=S) * P(I_2=F|W_2=R) = 0.2 * 0.8 = 0.16$
$w = P(I_1=T|W_1=R) * P(I_2=F|W_2=S) = 0.9 * 0.1 = 0.09$
$w = P(I_1=T|W_1=S) * P(I_2=F|W_2=S) = 0.2 * 0.1 = 0.02$

P(W₂ | I₁ = T, I₂ = F)

$$\hat{P}(W_2=R \mid I_1=T, I_2=F) = \frac{\hat{P}(W_2=R,\ I_1=T,\ I_2=F)}{\hat{P}(I_1=T,\ I_2=F)}$$
$$= \frac{0.72+0.16+0.72}{0.72+0.16+0.72+0.09+0.09+0.02} = 0.889$$

	0	1
A		x
B		x
C		x
D	x	
E		x

Weight = 0.48

	0	1
A		x
B		x
C		x
D	x	
E		x

Weight = 0.48

	0	1
A		x
B		x
C		x
D		x
E		x

Weight = 0.32

	0	1
A		x
B		x
C		x
D	x	
E		x

Weight = 0.48

	0	1
A		x
B		x
C	x	
D	x	
E		x

Weight = 0.64

$P(C=1|B=1, E=1).\ \frac{1.76}{2.4} = 0.733333333333$

Data를 선형으로 분리가 가능하다.

The data are linearly separable. **section_11**

Sec11_q3_3 : 새로 구한 최종 weight로 seperate 불가능하다
Points 3, 4에서 작동하지 않기 때문

$-1\cdot1 + 1\cdot2 + -1\cdot4 = -3 < 0$
$-1\cdot1 + 1\cdot3 + -1\cdot4 = -2 < 0$

$$w = w + y^* \cdot f$$

step	Weights	Score	Correct?
1	[-1, 0, 0]	$-1\cdot1+0\cdot1+0\cdot1=-1$	yes
2	[-1, 0, 0]	$-1\cdot1+0\cdot3+0\cdot2=-1$	no
3	[0, 3, 2]	$0\cdot1+3\cdot2+2\cdot4=14$	yes
4	[0, 3, 2]	$0\cdot1+3\cdot3+2\cdot4=17$	yes
5	[0, 3, 2]	$0\cdot1+3\cdot2+2\cdot3=12$	no

Final weights: [-1, 1, -1]

$$sample = R(s,a,s') + \gamma \max_{a'} Q(s',a')$$
$$Q(s,a) \leftarrow (1-\alpha)Q(s,a) + (\alpha)[sample]$$
$$V^{\pi}_{k+1}(s) \leftarrow \sum_{s'} T(s,\pi(s),s')[R(s,\pi(s),s') + \gamma V^{\pi}_k(s')]$$

Sample of V(s): $sample = R(s,\pi(s),s') + \gamma V^{\pi}(s')$

Update to V(s): $V^{\pi}(s) \leftarrow (1-\alpha)V^{\pi}(s) + (\alpha)sample$

Same update: $V^{\pi}(s) \leftarrow V^{\pi}(s) + \alpha(sample - V^{\pi}(s))$

$V^{\pi}(s) \leftarrow (1-\alpha)V^{\pi}(s) + \alpha\left[R(s,\pi(s),s') + \gamma V^{\pi}(s')\right]$

$Q_{k+1}(s,a) \leftarrow \sum_{s'} T(s,a,s')\left[R(s,a,s') + \gamma \max_{a'} Q_k(s',a')\right]$

$$\text{MEU}(e) = \max_a \sum_s P(s|e)\,U(s,a)$$

$$\text{VPI}(E'|e) = \left(\sum_{e'} P(e'|e)\text{MEU}(e,e')\right) - \text{MEU}(e)$$

- X is conditionally independent of Y given Z

if and only if: $X \perp\!\!\!\perp Y | Z$

$\forall x,y,z : P(x,y|z) = P(x|z)P(y|z)$

or, equivalently, if and only if

$\forall x,y,z : P(x|z,y) = P(x|z)$

$$P_{LAP,k}(x) = \frac{c(x)+k}{N+k|X|}$$

N : 관찰 데이터 개수
X 변수 종류의 개수
c(x) : 변수X의 관찰 데이터 개수

$$P_{LAP,k}(x|y) = \frac{c(x,y)+k}{c(y)+k|X|}$$

1. Calculate the expected net gain from buying car c, given no test.

$$EU(\text{buy}) = P(Q=+q)\cdot U(+q,\text{buy}) + P(Q=\neg q)\cdot U(\neg q, buy)$$
$$= .7\cdot500 + 0.3\cdot -200 = 290$$

2. Tests can be described by the probability that the car will pass or fail the test given that the car is in good or bad shape. We have the following information:

$$P(T=\text{pass}|Q=+q) = 0.9$$
$$P(T=\text{pass}|Q=\neg q) = 0.2$$

Calculate the probability that the car will pass (or fail) its test, and then the probability that it is in good (or bad) shape given each possible test outcome.

$$P(T=\text{pass}) = \sum_q P(T=\text{pass}, Q=q)$$
$$= P(T=\text{pass}|Q=+q)P(Q=+q) + P(T=\text{pass}|Q=\neg q)P(Q=\neg q)$$
$$= 0.69$$
$$P(T=fail) = 0.31$$
$$P(Q=+q|T=\text{pass}) = \frac{P(T=\text{pass}|Q=+q)P(Q=+q)}{P(T=\text{pass})} = \frac{0.9\cdot0.7}{0.69} = \frac{21}{23} \approx 0.91$$
$$P(Q=+q|T=\text{fail}) = \frac{P(T=\text{fail}|Q=+q)P(Q=+q)}{P(T=\text{fail})} = \frac{0.1\cdot0.7}{0.31} = \frac{7}{31} \approx 0.22$$

3. Calculate the optimal decisions given either a pass or a fail, and their expected utilities.

$$EU(\text{buy}|T=pass) = P(Q=+q|T=\text{pass})U(+q,\text{buy}) + P(Q=\neg q|T=\text{pass})U(\neg q,\text{buy})$$
$$\approx 0.91\cdot500 + 0.09\cdot(-200) \approx 437$$
$$EU(\text{buy}|T=\text{fail}) = P(Q=+q|T=\text{fail})U(+q,\text{buy}) + P(Q=\neg q|T=\text{fail})U(\neg q,\text{buy})$$
$$\approx 0.22\cdot500 + 0.78\cdot(-200) = -46$$
$$EU(\neg\text{buy}|T=pass) = 0$$
$$EU(\neg\text{buy}|T=fail) = 0$$

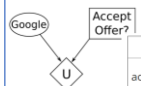
Therefore: $MEU(T=\text{pass}) = 437$ (with buy) and $MEU(T=\text{fail}) = 0$ (using ¬buy)

4. Calculate the value of (perfect) information of the test. Should the buyer pay for a test?

$$VPI(T) = \left(\sum_t P(T=t)MEU(T=t)\right) - MEU(\phi)$$
$$= 0.69\cdot437 + 0.31\cdot0 - 290 \approx 11.53$$

You shouldn't pay for it, since the cost is $50.

EU(accept) = P(hired)U(accept, hired) + P(not hired)U(accept, not hired) = 6500
EU(decline) = P(hired)U(decline, hired) + P(not hired)U(decline, not hired) = 2500

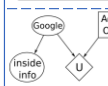


accept

Action	EU	
accept Acme offer		Answer: 6500.0
reject Acme offer		Answer: 2500.0

Action	Google outcome	U
accept Acme offer	hired	2000
accept Acme offer	not hired	8000
reject Acme offer	hired	10000
reject Acme offer	not hired	0

Google outcome	P(Google outcome)
hired	0.25
not hired	0.75



info	Google outcome	P(info | Google outcome)
good news	hired	0.7
bad news	hired	0.3
good news	not hired	0.1
bad news	not hired	0.9

info	P(info)
good news	0.25
bad news	0.75

Google outcome	info	P(Google outcome | info)
hired	good news	0.7
not hired	good news	0.3
hired	bad news	0.1
not hired	bad news	0.9

Info	Action	EU(Action | Info)	
good news	accept Acme offer		Answer: 3800.0
good news	reject Acme offer		Answer: 7000.0
bad news	accept Acme offer		Answer: 7400.0
bad news	reject Acme offer		Answer: 1000.0

A is your action, G is Google outcome, and I is info.
Use the following equation to calculate EU(a|i).
$EU(a|i) = \sum_G P(g|i)U(a,g)$
For example, here is how you would calculate EU(accept|good news).
EU(accept|good news) = P(hired|good news)U(accept, hired) + P(not hired|good news)U(accept, not hired) = 3800

EU(action|A) = SUM[P(★ |A) U(action, ★)]
(7000*0.25) + (7400*0.75)

If we are given the inside information, what is the expected value of MEU?
MEU(I) = P(good news)MEU(good news) + P(bad news)MEU(bad news) = 7300

What is the value of perfect information of the random variable Inside Info?
VPI(I) = MEU(I) - MEU(Ø) = 800

MEU(good news) = max(EU(accept|good news), EU(reject|good news)) = 7000

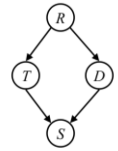
info	MEU(info)	
good news		Answer: 7000.0
bad news		Answer: 7400.0

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I'm sad
- Questions:

$T \perp\!\!\!\perp D$

$T \perp\!\!\!\perp D | R$ Yes

$T \perp\!\!\!\perp D | R, S$



Training Set

#	f_1	f_2	y^*
1	1	1	-
2	3	2	+
3	2	4	+
4	3	4	+
5	2	3	-

$$[w_0, w_1, w_2] = [-1, 0, 0]$$

Single Perceptron Update Pass

step	Weights	Score	Correct?	Update
1	[-1, 0, 0]	$-1\cdot1+0\cdot1+0\cdot1=-1$	yes	none
2	[-1, 0, 0]	$-1\cdot1+0\cdot3+0\cdot2=-1$	no	+[1, 3, 2]
3	[0, 3, 2]	$0\cdot1+3\cdot2+2\cdot4=14$	yes	none
4	[0, 3, 2]	$0\cdot1+3\cdot3+2\cdot4=17$	yes	none
5	[0, 3, 2]	$0\cdot1+3\cdot2+2\cdot3=12$	no	-[1, 2, 3]
6	[-1, 1, -1]			

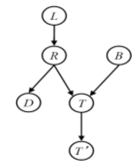
$$f(x) = \begin{bmatrix} -2 & 3 & 1 \end{bmatrix}$$

$\mathbf{w}_0 = [-2 \;\; 2 \;\; 1]$
$\mathbf{w}_1 = [0 \;\; 3 \;\; 4]$
$\mathbf{w}_2 = [1 \;\; 4 \;\; -2]$

$$\arg\max(Wx) = \arg\max\left(\begin{bmatrix} 11 \\ 13 \\ 8 \end{bmatrix}\right) = 1$$

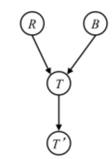
$w_1 = [0 \;\; 3 \;\; 4] - [-2 \;\; 3 \;\; 1] = [2 \;\; 0 \;\; 3]$

$w_2 = [1 \;\; 4 \;\; -2] + [-2 \;\; 3 \;\; 1] = [-1 \;\; 7 \;\; -1]$



$L \perp\!\!\!\perp T' | T$ Yes
$L \perp\!\!\!\perp B$ Yes
$L \perp\!\!\!\perp B | T$
$L \perp\!\!\!\perp B | T'$
$L \perp\!\!\!\perp B | T, R$ Yes

$R \perp\!\!\!\perp B$ Yes
$R \perp\!\!\!\perp B | T$
$R \perp\!\!\!\perp B | T'$



Conditional Probabilities

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

The Product Rule

$$P(y)P(x|y) = P(x,y) \Leftrightarrow P(x|y) = \frac{P(x,y)}{P(y)}$$

The Chain Rule

$$P(x_1,x_2,x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1,x_2)$$
$$P(X_1,X_2,\ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1,X_2)\ldots$$

Causal Chains

$$P(x,y,z) = P(x)P(y|x)P(z|y)$$

Common Cause

$$P(x,y,z) = P(y)P(x|y)P(z|y)$$

Bayes' Rule

$$P(x|y) = \frac{P(y|x)}{P(y)}P(x)$$

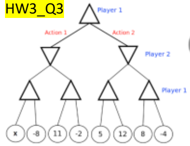
independent

$$P(x,y) = P(x)P(y) \dashrightarrow X \perp\!\!\!\perp Y$$

conditionally independent

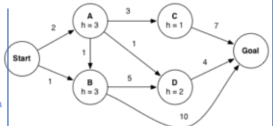
$$P(x,y|z) = P(x|z)P(y|z) \dashrightarrow X \perp\!\!\!\perp Y | Z$$

HW3_Q3

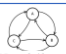


Q1)X>8 , Q3)None Q4)No

Q2) (x+11)/2 > 10 ⟶ x > 9



	Expand	return
A*	SBADCG	SADG
UCS	SBADCG	SADG



$V_{k+1}(A)$ -0.85 =1(-2 + 0.5(2.3)) HW4_Q3

$V_{k+1}(A) = max(Q_{k+1}(A, clockwise), Q_{k+1}(A, counterclockwise))$

$Q_{k+1}(A, clockwise) = T(A, clockwise, B)[R(A, clockwise, B) + \gamma V_k(B)] + T(A, clockwise, C)[R(A, clockwise, C) + \gamma V_k(C)]$

$Q_{k+1}(A, counterclockwise) = T(A, counterclockwise, B)[R(A, counterclockwise, B) + \gamma V_k(B)] + T(A, counterclockwise, C)[R(A, counterclockwise, C) + \gamma V_k(C)]$

Discount Factor, $\gamma = 0.5$

s	a	s'	T(s,a,s')	R(s,a,s')
A	Clockwise	B	1.0	-2.0
A	Counterclockwise	B	0.4	-1.0
A	Counterclockwise	C	0.6	-2.0
B	Clockwise	C	1.0	2.0
B	Counterclockwise	A	0.6	0.0
B	Counterclockwise	C	0.4	0.0
C	Clockwise	A	0.6	1.0
C	Clockwise	B	0.4	0.0
C	Counterclockwise	A	0.4	1.0
C	Counterclockwise	B	0.6	0.0

Part 2: Here is the formula to calculate $Q^*(A, clockwise)$.

$Q^*(A, clockwise) = T(A, clockwise, B)[R(A, clockwise, B) + \gamma V^*(B)] + T(A, clockwise, C)[R(A, clockwise, C) + \gamma V^*(C)]$

Part 3: Here is the formula to calculate $Q^*(A, counterclockwise)$.

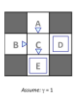
$Q^*(A, counterclockwise) = T(A, counterclockwise, B)[R(A, counterclockwise, B) + \gamma V^*(B)] + T(A, counterclockwise, C)[R(A, counterclockwise, C) + \gamma V^*(C)]$

Q^*(A, clockwise) -0.75 Q^*(A, counterclockwise) -0.8

optimal action from state A: Clockwise

HW5_Q3

Input Policy π



Observed Episodes (Training)

Episode 1: A, south, C, -1 / C, south, E, -1 / E, exit, x, +10

Episode 2: B, east, C, -1 / C, south, D, -1 / D, exit, x, -10

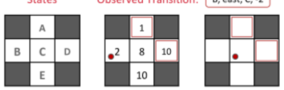
Episode 3: B, east, C, -1 / C, south, E, -1 / E, exit, x, +10

Episode 4: A, south, C, -1 / C, south, E, -1 / E, exit, x, +10

$\hat{V}^\pi(A)$	$\hat{V}^\pi(B)$	C	D	$\hat{V}^\pi(E)$
8	-2	4	-10	10

HW5_Q4

States Observed Transition: B, east, C, -2



$\hat{V}^\pi(A)$	$\hat{V}^\pi(B)$	C	D	$\hat{V}^\pi(E)$
1	4	8	10	10

$V^\pi(s) \leftarrow (1-\alpha)V^\pi(s) + \alpha\left[R(s,\pi(s),s') + \gamma V^\pi(s')\right]$

HW5_Q5 Model-Free RL: Cycle α γ 0.5.

	A	B	C
Clockwise	0.048	1.0	1.165
Counterclockwise	3.254	2.66	7.516

s	a	s'	r
A	Counterclockwise	C	1.0
C	Clockwise	A	0.0

답	A	B	C
Clockwise	0.048	1.0	1.584
Counterclo	4.006	2.660	7.516

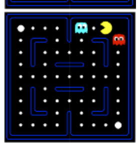
$Q(s,a) = (1-\alpha) * Q(s,a) + \alpha * (R(s,a,s') + \gamma * max_{a'} Q(s',a'))$

0.5 × 3.254 + 0.5 × (1 + 0.5 × 7.516)

Part1 1 번 답 $Q(s, West) = 1*1 + 10*3 = 31$

Part1 2 번 답 $Q(s, South) = 1*1 + 10*1 = 11$

Part1 3번 답 : 31 > 11 이므로 West 행동을 선택할 것이다.



Part2 1번 답 $Q(s', West) = 1*1 + 10*1 = 11$

Part2 2번 답 $Q(s', East) = 1*1 + 10*1 = 11$

Part2 3번 답 sample $= 9 + 1*11 = 20$

$\text{sample} = [r + \gamma \max_{a'} Q(s',a')] =$

Difference 구하기 $\text{difference} = [r + \gamma \max_{a'} Q(s',a')] - Q(s,a) =$

$\text{difference} = 20 - 31 = -11$

새로운 W1

$w_1 \leftarrow w_1 + \alpha\,(\text{difference})\, f_1(s,a) =$

$w_1 = 1 + .5*(-11)*1 = -4.5$

새로운 W2

$w_2 \leftarrow w_2 + \alpha\,(\text{difference})\, f_2(s,a) =$

$w_2 = 10 + .5*(-11)*3 = -6.5$

HW4_Q8 Discount Factor, $\gamma = 0.5$

s	a	s'	T(s,a,s')	R(s,a,s')
A	Clockwise	B	1.0	-1.0
A	Counterclockwise	B	0.2	-1.0
A	Counterclockwise	C	0.8	2.0
B	Clockwise	A	0.4	0.0
B	Clockwise	C	0.6	-2.0
B	Counterclockwise	A	1.0	-2.0
C	Clockwise	A	0.8	2.0
C	Clockwise	B	0.2	1.0
C	Counterclockwise	B	1.0	-2.0

$V^\pi_{k+1}(C)$ -2.65

$V^\pi_{k+1}(A) = T(A, clockwise, B)[R(A, clockwise, B) + \gamma V^\pi_k(B)] + T(A, clockwise, C)[R(A, clockwise, C) + \gamma V^\pi_k(C)]$

Q^π_∞(C, clockwise) 1.6

$Q^\pi_\infty(A, clockwise) = T(A, clockwise, B)[R(A, clockwise, B) + \gamma V^\pi_\infty(B)] + T(A, clockwise, C)[R(A, clockwise, C) + \gamma V^\pi_\infty(C)]$

Q^π_∞(C, counterclockwise) -3.0

$Q^\pi_\infty(A, counterclockwise) = T(A, counterclockwise, B)[R(A, counterclockwise, B) + \gamma V^\pi_\infty(B)] + T(A, counterclockwise, C)[R(A, counterclockwise, C) + \gamma V^\pi_\infty(C)]$

updated action for state C: Clockwise

HW5_Q2

s	a	s'	T(s,a,s')	R(s,a,s')
A	Clockwise	B	0.600	0
A	Clockwise	C	0.400	10
A	Counterclockwise	B	0.200	0.000
A	Counterclockwise	C	0.800	-1.000
B	Clockwise	A	0.400	-6.000
B	Clockwise	C	0.600	0.000
B	Counterclockwise	A	0.800	-4.000
B	Counterclockwise	C	0.200	0.000
C	Clockwise	A	0.600	0.000
C	Clockwise	B	0.400	-6.000
C	Counterclockwise	B	1.000	-9.000

$Q_k(s,a)$	A	B	C
Clockwise	2.8	-2.32	-1.68
Counterclockwise	-2.0	-1.84	-10.2

T(R + r*V()) + T(R + r*V():Q중에 큰거)

3.112=0.6(0+(0.5)*(-1.84))+0.4(10+(0.5)*(-1.68))

Q2 Part2 번 답

Q k+1(s,a)	A	B	C
Clockwise	3.112	-2.344	-1.928
Counterclo	-1.656	-2.248	-9.920

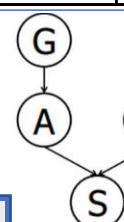
Q2 Part3 번 답

A	B	C
Clockwise	CounterCl	Clockwise

HW6_Q5

G	P(G)
g	0.10
¬g	0.90

G	A	P(A | G)
g	a	1.00
g	¬a	0.00
¬g	a	0.10
¬g	¬a	0.90



B	P(B)
b	0.40
¬b	0.60

A	B	S	P(S | A, B)
a	b	s	1.00
a	b	¬s	0.00
a	¬b	s	0.90
a	¬b	¬s	0.10
¬a	b	s	0.80
¬a	b	¬s	0.20
¬a	¬b	s	0.10
¬a	¬b	¬s	0.90

Q1 P(g, a, b, s) = P(g)*P(a|g)*P(b)*P(s|b, a) = 0.1*1*0.4*1 = 0.04

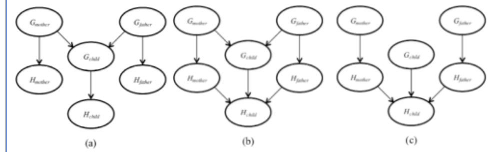
Q2 P(a) = P(a|g)*P(g) + P(a|¬g)*P(¬g) = 1.0*0.1+0.1*0.9 = 0.19

Q3 P(a|b) = P(a) = 0.19.

Q4 P(a|s, b) = P(a, b, s)/(P(a, b, s)+P(¬a, b, s)) = (P(a)*P(b)*P(s|a, b))/(P(a)*P(b)*P(s|a,b)+P(¬a)*P(b)*P(s|¬a, b)) = (0.19*0.4*1)/(0.19*0.4*1+0.81*0.4*0.8) = 0.076/(0.076+0.2592) = 0.2267

Q5 P(g|a) = P(g)*P(a|g)/(P(g)*P(a|g)+P(¬g)*P(a|¬g)) = 0.1*1/(0.1*1 + 0.9*0.1) = 0.1/(0.1+0.09) = 0.5263.

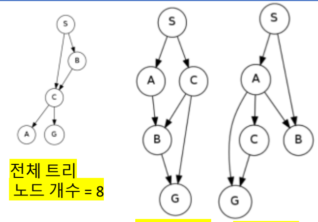
Q6 P(g|b) = P(g) = 0.1



$P(G_{father}, G_{mother}, G_{child}) = P(G_{father})P(G_{mother})P(G_{child})$ (c)

세 가지 네트웍 중에 독립성 가설과 일치한 것 = (a) (b)

가장 좋은 가설은 (a)

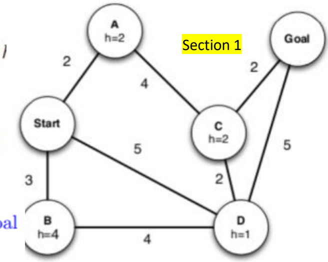


전체 트리 노드 개수 = 8

DFS = SABG

BFS = SAG

Section 1



d) Greedy search with the heuristic h.
States Expanded: Start, D, Goal
Path Returned: Start-D-Goal

e) A* search with the same heuristic.
States Expanded: Start, A, D, B, C, Goal
Path Returned: Start-A-C-Goal

a) Depth-first search.
States Expanded: Start, A, C, D, B, Goal
Path Returned: Start-A-C-D-Goal

b) Breadth-first search.
States Expanded: Start, A, B, D, C, Goal
Path Returned: Start-D-Goal

c) Uniform cost search.
States Expanded: Start, A, B, D, C, Goal
Path Returned: Start-A-C-Goal