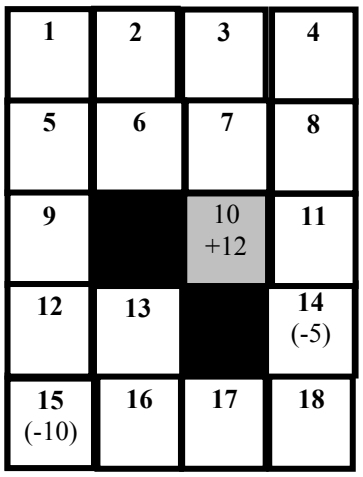
1. IGAZ/HAMIS?
   a. Az állapot hasznosságát csak akkor tudjuk értelmezni, ha az adott állapotból csak egy végállapotba lehet eljutni.
   b. A szekvenciális probléma iteratív megoldása során gyakran az értékek még nem konvergálnak pontosan, de a stratégia már egyértelmű lehet.
   c. Mivel az időbeli különbség tanuláshoz nincs szükség az állapotátmenet-modellre, ezért Q-tanulásra is használható.
   d. A hasznosságfüggvény explicit reprezentációja jobb általánosító képességet tesz lehetővé, mint az implicit reprezentáció.
   e. Az adott állapothoz tartozó legnagyobb Q érték az állapot hasznosságát adja.
   f. A mohó felfedezést végző ágens könnyen egy szuboptimumba juthat.
   g. A hóbortos felfedezést végző ágens nagyon jól kiismeri környezetét, de nem aknázza ki ezt a tudását a jutalmak gyűjtésére.
   h. Az aktív megerősítéses tanulásban kettős célja van az ágensnek: egyrészt minél nagyobb jutalmakat akar gyűjteni, másrészt javítani akarja a megszerzett tudást, a jövőbeli szekvenciákhoz.
   i. Eljárásmód-iteráció esetén, ha a *t*-dik lépésben kiszámítottuk az $U_t(s)$ értékeket, akkor a következő eljárásmód becslés – $\pi_{t+1}(s)$ – nem függ a leszámítolási tényezőtől.

2. Ha egy adott állapotból p_1 valószínűséggel jutunk 3 lépésben a V_1 végállapotba, és ez alatt J_1 jutalmat gyűjtünk, illetve p_2 valószínűséggel jutunk 8 lépésben a V_2 végállapotba, és ez alatt J_2 jutalmat gyűjtünk, akkor mekkora lesz a vizsgált állapot hasznossága? (A leszámítolási tényező 1.)

3. Az alábbi útvesztőben passzív megerősítéses tanulást végzünk. Az egyes mezőkben a felső szám az állapot sorszáma, az alsó, zárójelben lévő szám – ha van – az állapothoz rendelt jutalom értéke. Ahol nincs feltüntetve a jutalom értéke, ott nulla. A végállapot a szürke színnel is megjelölt 10-es.



Időbeli különbség tanulást végzünk az alábbi lépéssorozat mentén, a tanulás bátorsági faktora 0,1;a leszámítolási tényező 0,8:

$$1 \Rightarrow 5 \Rightarrow 9 \Rightarrow 12 \Rightarrow 13 \Rightarrow 16 \Rightarrow 17 \Rightarrow 18 \Rightarrow 14 \Rightarrow 11 \Rightarrow 8 \Rightarrow 11 \Rightarrow 14 \Rightarrow 11 \Rightarrow 8 \Rightarrow 7 \Rightarrow 10$$

Mi lesz a 11-es állapot hasznosságértékének új becslése a lépéssorozat után, ha a lépéssorozatot megelőzően a hasznosságbecslések a következők voltak:

U1 = 5	**U2 = 6**	**U3 = 7**	**U4 = 8**	**U5 = 7**	**U6 = 8**	**U7 = 9**	**U8 = 8**	**U9 = -6**
U10 = +12	**U11 = +10**	**U12 = -9**	**U13 = -9**	**U14 = -4**	**U15 = -9**	**U16 = -7**	**U17 = -7**	**U18 = -7**

(Ne törődjön azzal, hogy kialakulhatott-e ez a hasznosság becslés!)

4. Mi az ε-mohó felfedezési stratégia célja és lényege?
5. Boltzmann felfedezési stratégiát alkalmazunk. Egy *s* állapotban a lehetséges cselekvések hasznossága Q(a1,s)=1; Q(a2,s)=2; Q(a3,s)=1. Milyen valószínűséggel választjuk az a1, a2 és a3 cselekvést, ha a „hőmérséklet” $T=10^6$, ha T=1, illetve ha $T=10^{-6}$?
6. Egy aktív megerősítéses tanulási problémában 3 állapot alkotja az állapotteret, S3 a végállapot. Minden állapotban kétféle cselekvés (a1 és a2) közt választhatunk. Az alábbi baloldali táblázatban láthatók az a1-hez tartozó állapotátmenet-valószínűségek, a jobboldali táblázatban az a2 cselekvéshez tartozók. A leszámítolási tényező γ=0,5.

a_1 esetén $T(s \to s')$		*s'*		
		S1	S2	S3
s	S1	0,5	0	0,5
	S2	0	0,5	0,5
	S3	0	0	0

a_2 esetén $T(s \to s')$		*s'*		
		S1	S2	S3
s	S1	0	0,5	0,5
	S2	0,2	0	0,8
	S3	0	0	0

Az egyes állapotokhoz tartozó jutalmak $R(S_1) = -0,5$; $R(S_2) = +0,4$; $R(S_3)=+1,2$. Eljárásmód-iterációs algoritmust alkalmazunk, az eddigi iterációk eredményeképp a *t*. lépésben a becsült eljárásmód $\pi_t(S_1) = a2$; $\pi_t(S_2) = a2$, a becsült hasznosságok $U_t(S_1) = 0$; $U_t(S_2) = 1$; $U_t(S_3)= R(S_3)= 1,2$. (Ne törődjön vele, hogy kialakulhattak-e ezek az értékek!)

A következő iterációs lépés végére milyen eljárásmódot és milyen hasznosságértékeket kapunk?