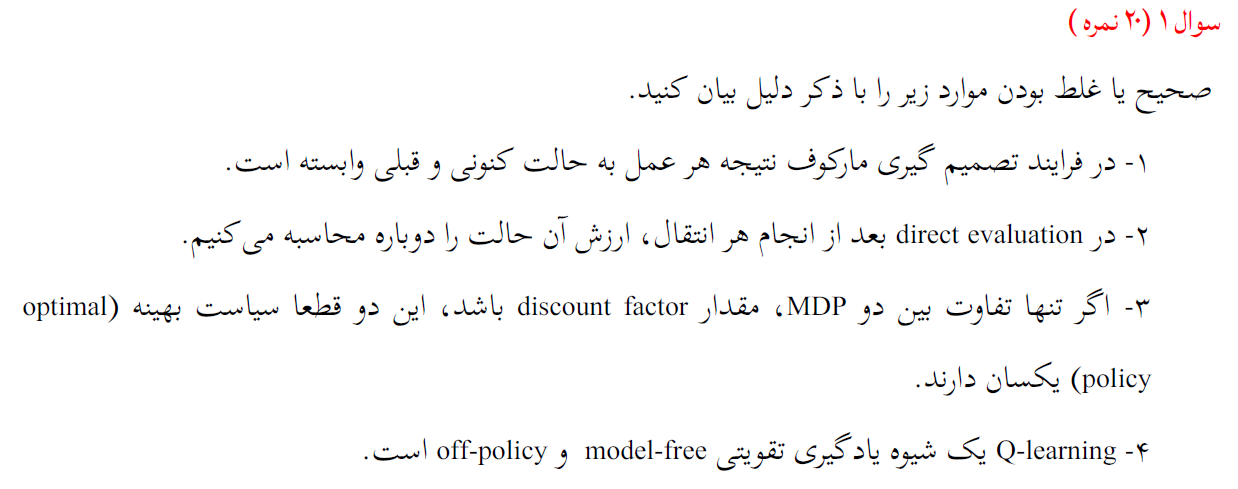
**باسمه تعالی**

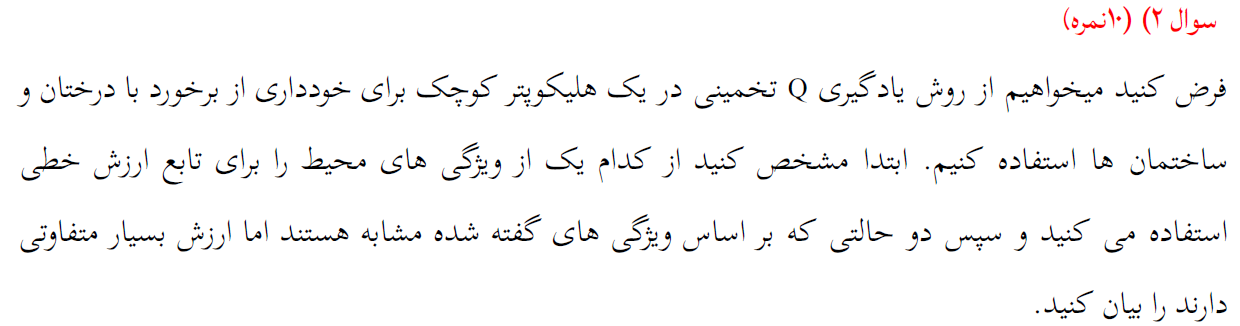
**تکلیف 3 درس هوش مصنوعی – دکتر جوانمردی**

**سید امیرمهدی میرشریفی - 9831105**

****

پاسخ:

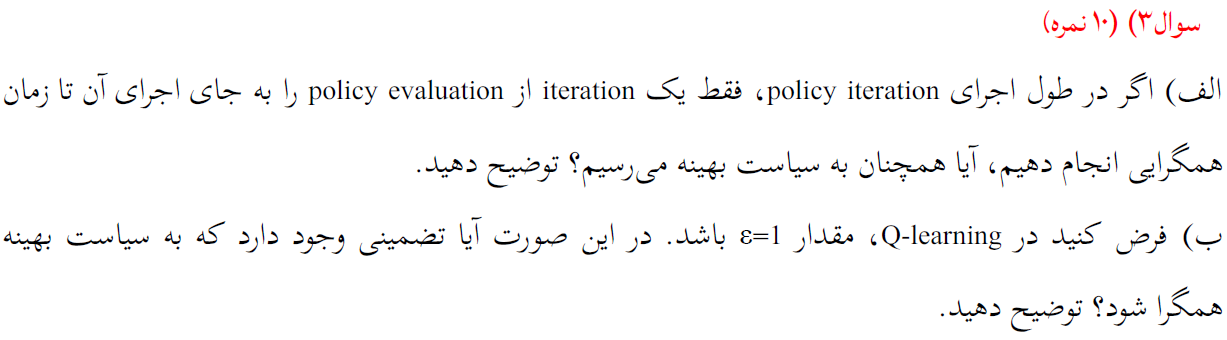
1. غلط - در تصمیم گیری مارکوف نتیجه هر عمل ما صرفا به حالت کنونی وابسته است.
2. غلط- اگر مقصود جمله این است که بعد از هر نمونه نتیجه را به روز رسانی میکنیم ، جواب غلط است در این روش ما در تمام بخش های نمونه مان ، دنباله هایی که شروعشان state مدنظرمان هست را به عنوان یک نمونه در نظر میگیریم و ارزش آن را با میانگیری آن ها حساب می کنیم.
3. غلط – در مسئله خوردن دات ها این قضیه نقض خواهد شد.
4. درست



پاسخ:

ویژگی ها همچون ، ارتفاع ، فاصله تا درخت ، سرعت ، جهت ، تراکم درختان اطراف و ....

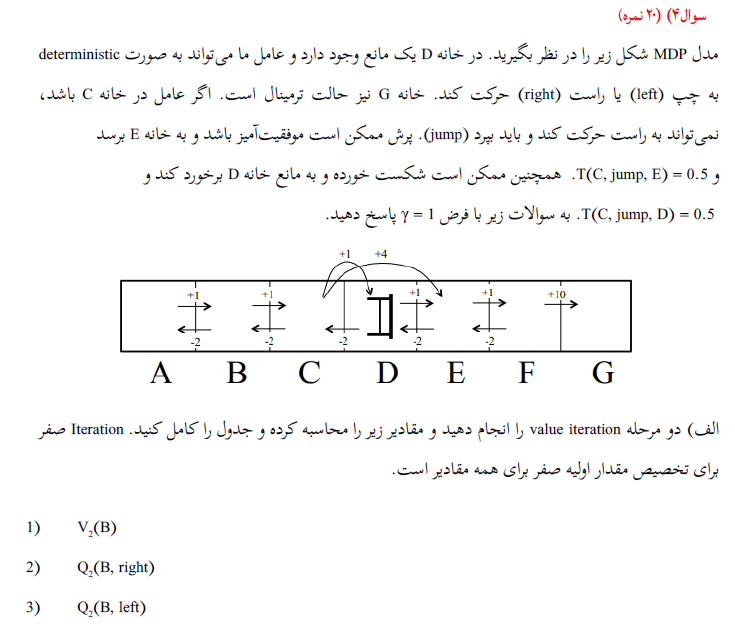
برای ساخت تابع ارزش خطی میتوان از پارامتر های سرعت( عامل خنثی یا مثبت) ، پراکندگی (عامل منفی) و فاصله تا درخت( عامل منفی) استفاده کرد. اگر هلی کوپتر در حالتی باشد که پراکندگی محیط زیاد است اما فاصله اش تا نزدیک ترین درخت متوسط باشد ارزش آن برابر خواهد شد با حالتی که پراکندگی کم است اما فاصله اش تا نزدیک ترین درخت بسیار بسیار کم باشد. به این صورت شرایط متفاوت اما ارزش برآورد شده شان یکی خواهد بود.



پاسخ:

الف) وقتی یک دور را بخواهیم مدام اجرا کنیم یعنی با یک سیاست یکسان هر دفعه مقادیر حالات را حساب کنیم ، بنابراین بعد از دور اول که مقادیر با سیاست ها همگرا شد سراغ دور دوم میرویم با همان سیاست ها که طبق همگرایی مرحله آخر مجددا همگرا است بنابراین لزوما بهینه نیست.

ب) وقتی اپسیلون برابر یک باشد یعنی همیشه رندوم عمل کند و در نهایت که همگرا میشود ، بهینه هم هست



پاسخ:

**In v1:**

**Va= vb +1= 0+1= 1**

**Vb= max ((vc +1) , va -2) = max(1,-1)= 1**

**Vc= max ((vd +1)/2+( ve +4)/2, -2+1) = max(5/2,2,-1)=5/ 2**

**Vd= max( vc -2 , ve +1 )= max(0/5, 1)=1**

**Ve= max ((vf +1) , vd -2) = max(1,-2)= 1**

**Vf= max ((vg +10) , ve -2) = max(10,-1)= 10**

**In v2:**

**Va= vb +1= 1+1= 2**

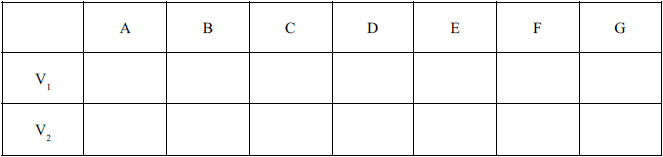
**Vb= max ((vc +1) , va -2) = max(3.5,0)= 3.5**

**Vc= max ((vd +1)/2+( ve +4)/2, -2+3.5) = max(7/2,2,1.5)=7/ 2**

**Vd= max( vc -2 , ve +1 )= max(1.5, 2)=2**

**Ve= max ((vf +1) , vd -2) = max(11,0)= 11**

**Vf= max ((vg +10) , ve -2) = max(10,-9)= 10 => V(B)=3.5 / Q(B,left)=0 / Q(B,right) =3.5**



**5/2**

**0**

**10**

**1**

**1**

**1**

**1**

**2**

**7/2**

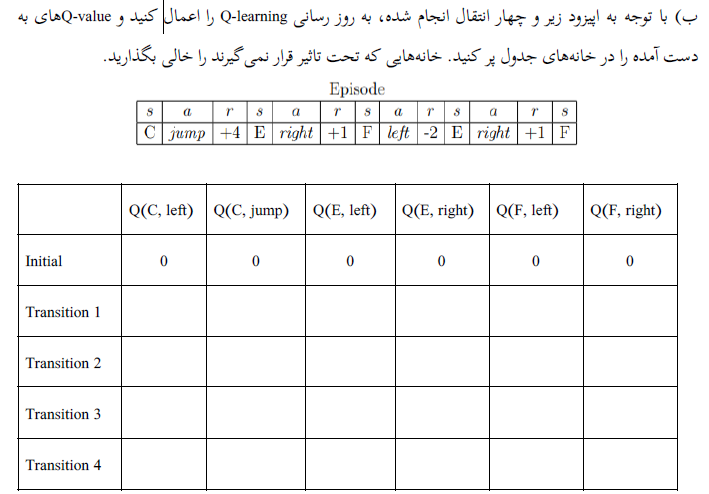
**2**

**3.5**

**11**

**10**

**0**



**0**

**1-**

**75/**

**0**

**2**

**0**

**0**

**2**

**0**

**5/**

**0**

**0**

**بر فرض آن که آلفا برابر نیم باشد:**

**0**

**5/**

**0**

**0**

**2**

**0**

**0**

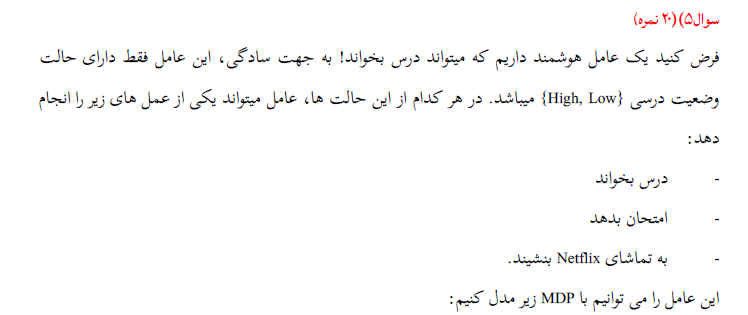
**0**

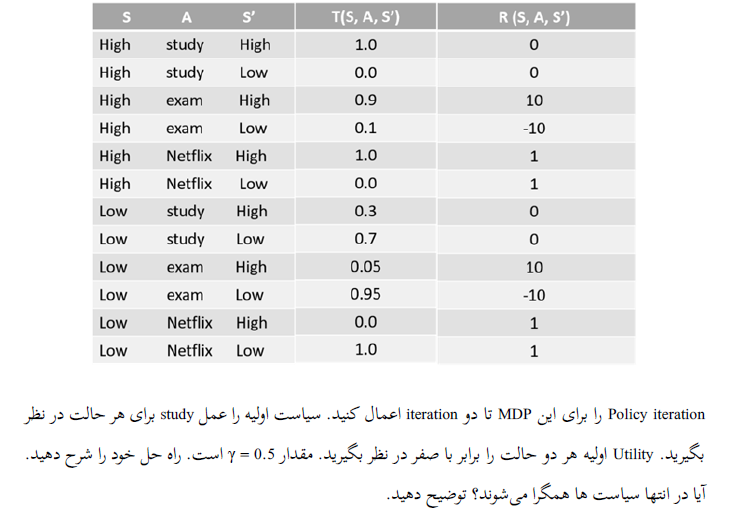
**0**

**0**

**2**

**0.75**-





پاسخ:

Π0=study and v1 =0 and v2 =0

V1 = 1\*(0+0) + 0\*(0 + 0)=0

V2 = .3\*(0+0) + 0\*(0+0) =0

* Now check Q for Each V :

Q(V1,study) = 0

Q(V1,exam)=.9\*(10+0) + .1\*(-10+0)=8 => best action = exam

Q(V1,Netflix)=1+0=1

Q(V2,study) = 0

Q(V2,exam)=.05\*(10+0) + .95\*(-10+0)=-9 => best action = netflix

Q(V1,Netflix)=1+0=1

* Π1 = exam, netflix

V1 = 8

V2 = 1+0=1

* Now check Q for Each V :

Q(V1,study) = 1\*(0 +8/2)=4

Q(V1,exam)=.9\*(10+8/2) + .1\*(-10+1/2)= 12.6 -.95 = 11.65 => best action=exam

Q(V1,Netflix)=1+8/2=5

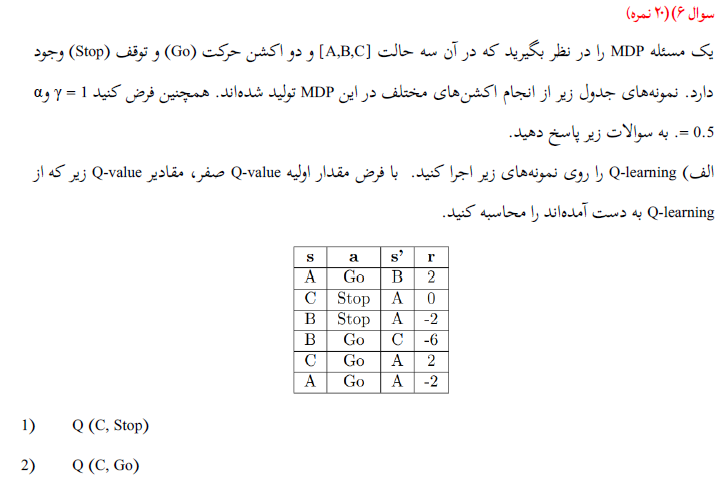
Q(V2,study) = .3\*(8/2) + .7(1/2)=1.2 + .35 = -1.55

Q(V2,exam)=.05\*(10+8/2) + .95\*(-10+1/2)=.7 – 9.025=-8.325 => best action= study

Q(V1,Netflix)=1-1/2=-.5

* Π2= for v1 = exam & for v2 = study

برای حل این مسئله همانطور که در بالا نوشته شده است اول با سیاست این که هر دو حالت درس جلو میرویم. ارزش حالت ها را به دست می آوریم. سپس بهترین اکشن را به عنوان سیاست بعدی پیدا میکنیم و همین مراحل را دو بار انجام میدهیم تا در نهایت برای حالت یک سیاست آزمون و حالت دو سیاست درس انتخاب میشود. برای این که بفهمیم همگرا شده است یا نه نیاز به یک مرحله دیگر دارد زیرا تا آخرین مرحله همگرا نبود.



پاسخ:

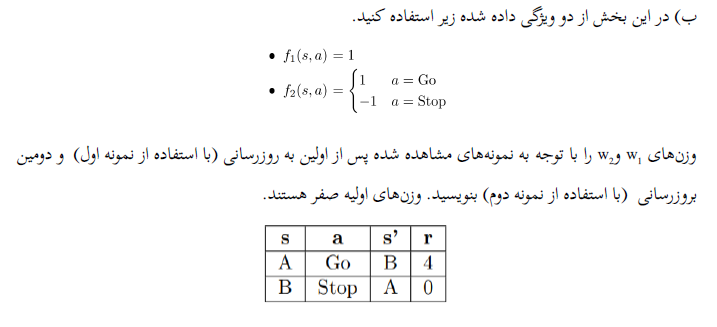
First (A,Go,B,2) => Q(A,Go) += .5 \* (2 + max Q(B) – Q(A,Go)) => Q(A,GO)= 1

Then (C,stop,A,0) => Q(C,stop) += .5\*(0 + maxQ(A) – Q(C,stop)) => Q(C,stop)=0.5

Then (B,stop,A,-2) => Q(B,stop) += .5\* ( -2 + maxQ(A) - Q(B,stop))=> Q(B,stop) =-0.5

Then (B,Go,C,-6) => Q(B,Go)+=0.5\*(-6 + maxQ(c)- Q(B,Go)) => Q(B,Go)=-0.275

Then (C ,GO ,A ,2)=> Q(C,Go)+=0.5\*(2+maxQ(A)- Q(C,Go))=> Q(C,Go)=1.5

Finally (A,GO,A,-2)=> Q(A,Go) += 0.5\*(-2 + maxQ(A)- Q(A,Go)) => Q(A,Go)=0

پاسخ:

با توجه به آن که آلفا بیان نشده است مقدار آن را نیم در نظر میگیریم:

Q(S,A) = w1\*f1(s,a) + w2\*f2(s,a)

W1=w2=0

First: Q(A,Go) = 0 + 0 = 0

R + maxQ(B) – Q(A,Go) = 4 + 0 – 0=0 -> w1 = w1 + alpha\*difference\*fi = 0 + .5\* 4 \*1 =2

W2 = w2 + alpha\*difference\*f2 = 0 + .5\*1 = 2

W1 = W2 = 2

Then:

Q(B,Stop) = 2\* 1 + 2\*-1= 0

Difference = 0 + maxQ(A) – Q(B,stop) = 0 + 3 – 0 = 3 => w1 = 2 + 0.5\*3\*1 = 3.5

W2= 2 + .5\*3 \* (-1) = 0.5

W1 = 3.5 W2 = 0.5