NAME: PAURAVI UDAY NAGARKAR.

ID: W1650209

Honework No: 5

Broken 2 Game has two States Play & Win. There are three actions available at State Play; Play the Blue machine, Green machine, Red machine. The account represent the State teansitions, and actions are labelled on accounts. however four Playing Blue & Red machine, the teansition Probabilities are unknown and Once were State is reached, game is over.

win

play
greem
with Probably
of 1.0

The reward are given;

R (5, 9,8') a S green Play Play Play Blue Play 10 Play win Play Play Red Z'a Wing 50 Play Red

use temposal difference learning to learn the value of Play State, by successively, applying the two episode below. The initial values of all States from 0.

with learning mate d = 0.5 and discount factor

y = 0.5 what value of State Play do we learn?

	" (W.) was
	s a S'
	Play Green Play
11111	Play Red Play
2.3	The state of the s
	(Play, Blue, Play)
-	No.
A A	Play, (Play) (Win)
1, M. M. J.	Play)
	Red LPlay, Red, Win,
•	(Play, Red, Play)
	riag, rear ray
1	(Time)
	00h d - 0.5
-	Learning Rate d = 0.5
* A	Discount factor (= 0.5 Episodes Given: - (S, 9, S') - (Play, Green, Play)
-	(Play, Red, Play)
	as for the pure of
	using temporal difference learning to learn the value
	using temporar augustice rearrang to
	as the 'play state'
	updated V(S) + V(S) + V[Sample - V(S)] 4
	Sample = $R(S, a, S') + YV(S!)$
	my Sample = MCS, 4, 5 miles Sony
	Tout always the state's value as a
	Initializing the State's value as O Vo(S) = 0, for S = Play win
7 1	Eirst Episode
	Sample = R (Play, Green, Play) + YV (Play)
1 61 - 1	$= 2 + 0.5 \pm 0 = 2$
	updated Value
(UM-)	v(Play) = v(Play) + & (Sample - v(play))
Vacanta de la constanta de la	= 0 -1 0.5[2-0]
19	
	V (Play) = 1 aplex Episade 1.

(M. Irally In.

\sim				
	Second Episode			
	Sample	= R(play, Red	Play) + YV(1	Play)
,		= 0+0,5×11	= 0.5	
1		9		7
	V (Play) =	V(Play) + &	Sample - U(Pl	ay 1
	=	1 + 0.5 x	(0.5-1)	
La Maria de la Maria della Mar	y (Play) =	0.75 Capter	e 2 nd episocle)	
	Updated V(Play) = V(Play) + 2[Sample - V(Play)] = 1 + 0.5 * (0.5 - 1) V(Play) = 0.75 (after 2 nd episorele)			
	Value of Sto	ile Play for Episo	ado 1 : 1	
132 1 2 3	value of St	re Play for Episo	20de 2: 0:75	-
	,		1 1 2 2	
Problem 2	for Same	game shown i	n figure 1, w	<u>e</u>
	<u>apseeved</u>	three episodes		
	S	a Oneen	Olau	
	Play	Oneen	Play	
	Play	Red	win	
	0	Blue	Ĉ	
	0 1	ung to update the	he values of O-	State
· · · · · · · · · · · · · · · · · · ·	*			
· fund	by apprying	a learning ab	0.5 and a disc	ount of
	anomie use	ize all Q states vi	alues Os O	V.
1	U.S. SINCLE	ige tour of the second		
	S	a	9(5,0)	
(1)	Play	Green	i - Barry	
6	Play	Red		
	Play	Blue	10000	_
	7.00	7. K	-	y
			A series	<u> </u>
	- Almandal 4	1 May make the		
1		V: 1 4 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	2 - 3 A	6
		18 m	1 1 1	Y
	R .			6

```
learning Rate d = 0.5
                     (S, a, s') -> (Play, Green, Play)
[Play, Red, Play
 Discount factor 7 = 0.5
   Episodes given
                                (play, Blue, win)
using g karening to update the values of g. States

Intializing g States values as 101

Op (S) = 0 for S = Play, win.
updated g(s,a) - g(s,a) + d. (sample - g(s,a)
  Sample = R(S, 9, S!) + y Max (g'(S', 9')
 where
     Sample = R[play, green, Play) + 0.5 * Max
  Episode 1:
              ( 'g ( Play a'))
    = 2 + 0 + 0.5
 updated Play I have the
     Q(S,Q)=0+0.5(2-0)
g (Play, green) = 1111
Sample = R(Play, Red. Play) + 0.5* max (g(Play, a'))
   Sample = 0 + 0.5 * 1
 updated
   Q(S,Q) = Q(Play, Red) + 0.5 + (Sample - Q(Play, Red)
          = 0+0.5 * (0.5-0)
   Q(S,a) = 0.25
```

,					
	g (play, Red) = 0.25				
	Episoele 3 Sample = R(Play, Blue, Win) + Y Max Q (win, a')				
	= 10+0.5*0				
	= = = 10				
	updated				
	updated g(s,q) = g(Play, Blue) = g(Play, Blue) + x (Sample - g(Play, Blue) - 0 + 0.5 * (10-0) a(Play, Blue) = 5				
	Q(Play, Blue) = 5				
	5 a g(S,a)				
	Play green 1 Play Red 0.25 Play Blue 5				
will a second					
	71.7				
10 M					
,					
1					