Question 1

Having the following XO sequence of states a long with their values

X - O O - X X - O	0.500
X - O O X X X - O	0.537
X O O O X X X - O	0.577
X O O O X X X X O	0.500

Assume a learning rate of 0.69 what will be updated values adopting gradient-based state value update with each move.
Answer:
$v(t)=v(t)+\eta*(v(t+1)-v(t))$
The updated values are:
0.525368
0.564661
0.523930
0.500000

Question 2

Having the following XO sequence of states a long with their values

O O X - - X - - -	0.500
O O X - - X - X -	0.543
O O X - - X O X -	0.506

O O X X - X O X -	0.570
O O X X O X O X -	0.509
O O X X O X O X X	1.000

Assume a learning rate (η) of 0.50 and discount factor (γ) of 0.88, what will be updated values adopting TD-based state value update with each move. Assume all rewards are -1 except for the actions leading to the goal state with respect to X-player.
Answer:
v(t)=v(t)+eta*(r+gama*v(t+1)-v(t))
The updated values are:
-0.011099
-0.005908
0.003613
0.008885
0.694618
1.000000

Question 3

Having the following XO sequence of states a long with their values

O - X - - X - - O	0.500
O X X - - X - - O	0.598
O X X - - X O - O	0.515
O X X - X X O - O	0.576
O X X - X X	0.000

O	O	O	

Assume a learning rate of 0.84 what will be updated values adopting gradient-based state value update with each move.
Answer:
v(t)=v(t)+η*(v(t+1)-v(t))
The updated values are:
0.582558
0.528692
0.566151
0.092130
0.000000

Question 4

Having the following XO sequence of states a long with their values

X	O	O	0.500
-	-	-	
O	X	X	
X	O	O	0.567
X	-	-	
O	X	X	
X	O	O	0.542
X	-	O	
O	X	X	
X	O	O	1.000
X	X	O	
O	X	X	

Assume a learning rate (η) of 0.70 and discount factor (γ) of 0.96, what will be updated values adopting TD-based state value update with each move. Assume all rewards are -1 except for the actions leading to the goal state with respect to X-player.
Answer:
v(t)=v(t)+eta*(r+gama*v(t+1)-v(t))
The updated values are:
-0.168895
-0.165704
0.834571
1.000000

Question 5

Having the following XO sequence of states a long with their values

- O - O X - X X O	0.500
X O - O X - X X O	0.509
X O - O X O X X O	0.507
X O X O X O X X O	1.000

Assume a learning rate of 0.97 what will be updated values adopting gradient-based state value update with each move.
Answer:
$v(t)=v(t)+\eta*(v(t+1)-v(t))$
The updated values are:
0.508446
0.506755
0.985201
1.000000

Question 6

Having the following XO sequence of states a long with their values

- X O O O - X - X	0.500
X X O O O - X - X	0.549
X X O O O - X O X	0.510
X X O O O X X O X	0.500

Assume a learning rate (η) of 0.78 and discount factor (γ) of 0.95, what will be updated values adopting TD-based state value update with each move. Assume all rewards are -1 except for the actions leading to the goal state with respect to X-player.
Answer:
v(t)=v(t)+eta*(r+gama*v(t+1)-v(t))
The updated values are:
-0.263209
-0.281383
-0.297320
0.500000

Question 7

Having the following XO sequence of states a long with their values

X - - - O X X O O	0.500
X - X - O X X O O	0.543
X - X O O X X O O	0.527
X X X O O X X O O	1.000

Assume a learning rate of 0.87 what will be updated values adopting gradient-based state value update with each move.
Answer:
v(t)=v(t)+η*(v(t+1)-v(t))
The updated values are:
0.537694
0.529043
0.938498
1.000000

Question 8

Having the following XO sequence of states a long with their values

X - - - O -	0.500

- X O	
X - X - O - - X O	0.521
X O X - O - - X O	0.541
X O X - O X - X O	0.581
X O X O O X - X O	0.542
X O X O O X X X O	0.500

Assume a learning rate (η) of 0.99 and discount factor (γ) of 0.97, what will be updated values adopting TD-based state value update with each move. Assume all rewards are -1 except for the actions leading to the goal state with respect to X-player.
Answer:
v(t)=v(t)+eta*(r+gama*v(t+1)-v(t))
The updated values are:
-0.484993
-0.465585
-0.426568
-0.463990
-0.504433
0.500000

Question 9

Having the following XO sequence of states a long with their values

O - - - X X O X O	0.500
O - X - X X O X O	0.528
O O X	0.512

- X X O X O	
O O X X X X O X O	1.000

Assume a learning rate of 0.70 what will be updated values adopting gradient-based state value update with each move.
Answer:
v(t)=v(t)+η*(v(t+1)-v(t))
The updated values are:
0.519947
0.517146
0.853684
1.000000

Question 10

Having the following XO sequence of states a long with their values

O O - X X - - X O	0.500
O O X X X - - X O	0.510
O O X X X O - X O	0.552
O O X X X O X X O	1.000

Assume a learning rate (η) of 0.57 and discount factor (γ) of 0.91, what will be updated values adopting TD-based state value update with each move. Assume all rewards are -1 except for the actions leading to the goal state with respect to X-player.
Answer:
v(t)=v(t)+eta*(r+gama*v(t+1)-v(t))
The updated values are:
-0.090582
-0.064632
0.755931
1.000000

Question 11

Having the following XO sequence of states a long with their values

X O - - O X - X O	0.500
X O - - O X X X O	0.558
X O O - O X X X O	0.511
X O O X O X X X O	1.000

Assume a learning rate of 0.72 what will be updated values adopting gradient-based state value update with each move.
Answer:
$v(t)=v(t)+\eta*(v(t+1)-v(t))$
The updated values are:
0.541682
0.524318
0.863153
1.000000

Question 12

Having the following XO sequence of states a long with their values

X O X - O X - - O	0.500
X O X X O X - - O	0.519
X O X X O X O - O	0.555
X O X X O X	0.500

O	X	O	

Assume a learning rate (η) of 0.78 and discount factor (γ) of 0.85, what will be updated values adopting TD-based state value update with each move. Assume all rewards are -1 except for the actions leading to the goal state with respect to X-player.
Answer:
v(t)=v(t)+eta*(r+gama*v(t+1)-v(t))
The updated values are:
-0.326055
-0.298000
-0.326432
0.500000

Question 13

Having the following XO sequence of states a long with their values

O	O	X	0.500
-	-	X	
-	X	O	
O	O	X	0.544
-	-	X	
X	X	O	
O	O	X	0.576
O	-	X	
X	X	O	
O	O	X	1.000
O	X	X	
X	X	O	

Assume a learning rate of 0.88 what will be updated values adopting gradient-based state value update with each move.
Answer:
v(t)=v(t)+η*(v(t+1)-v(t))
The updated values are:
0.538635
0.572285
0.949139
1.000000

Question 14

Having the following XO sequence of states a long with their values

X	-	X	0.500

O - O - - -	
X - X O - O - - X	0.597
X - X O - O O - X	0.558
X - X O - O O X X	0.565
X - X O O O O X X	0.000

Assume a learning rate (η) of 0.79 and discount factor (γ) of 0.82, what will be updated values adopting TD-based state value update with each move. Assume all rewards are -1 except for the actions leading to the goal state with respect to X-player.
Answer:
v(t)=v(t)+eta*(r+gama*v(t+1)-v(t))
The updated values are:
-0.298166
-0.302966
-0.306758
-0.671349
0.000000

Question 15

Having the following XO sequence of states a long with their values

- O - O - X X X O	0.500
X O - O - X X X O	0.556
X O O O - X X X O	0.586
X O O	0.500

O X X X X O	

Assume a learning rate of 0.66 what will be updated values adopting gradient-based state value update with each move.
Answer:
v(t)=v(t)+η*(v(t+1)-v(t))
The updated values are:
0.536733
0.575601
0.529198
0.500000

Question 16

Having the following XO sequence of states a long with their values

- X - - - - O O X	0.500
- X - X - - O O X	0.585
- X - X - O O O X	0.526
X X - X - O O O X	0.558
X X O X - O O O X	0.558
X X O X X O O O X	1.000

Assume a learning rate (η) of 0.64 and discount factor (γ) of 0.97, what will be updated values adopting TD-based state value update with each move. Assume all rewards are -1 except for the actions leading to the goal state with respect to X-player.
Answer:
v(t)=v(t)+eta*(r+gama*v(t+1)-v(t))
The updated values are:
-0.096810

-0.103134
-0.104107
-0.092256
0.821847
1.000000

Question 17

Having the following XO sequence of states a long with their values

- O X X X O - - O	0.500
- O X X X O - X O	0.518
O O X X X O - X O	0.599
O O X X X O X X O	1.000

Assume a learning rate of 0.93 what will be updated values adopting gradient-based state value update with each move.
Answer:
$v(t)=v(t)+\eta*(v(t+1)-v(t))$
The updated values are:
0.517097
0.593752
0.971960
1.000000

Question 18

Having the following XO sequence of states a long with their values

- - X O - - - X O	0.500
- - X O X - - X O	0.600

- - X O X O - X O	0.596
X - X O X O - X O	0.530
X O X O X O - X O	0.590
X O X O X O X X O	1.000

Assume a learning rate (η) of 0.77 and discount factor (γ) of 0.89, what will be updated values adopting TD-based state value update with each move. Assume all rewards are -1 except for the actions leading to the goal state with respect to X-player.
Answer:
v(t)=v(t)+eta*(r+gama*v(t+1)-v(t))
The updated values are:
-0.244130
-0.223400
-0.269324
-0.243590
0.821028
1.000000

Question 19

Having the following XO sequence of states a long with their values

- - - X O O - X -	0.500
X - - X O O - X -	0.528
X - - X O O O X -	0.533
X - - X O O	0.540

O X X	
X O - X O O O X X	0.541
X O X X O O O X X	0.500

Assume a learning rate of 0.81 what will be updated values adopting gradient-based state value update with each move.
Answer:
$v(t)=v(t)+\eta*(v(t+1)-v(t))$
The updated values are:
0.522582
0.531771
0.538931
0.540679
0.507742
0.500000

Question 20

Having the following XO sequence of states a long with their values

X X - O - - - - O	0.500
X X - O - - X - O	0.503
X X - O O - X - O	0.513
X X - O O - X X O	0.535
X X - O O O X X O	0.000

Assume a learning rate (η) of 0.72 and discount factor (γ) of 0.81, what will be updated values adopting TD-based state value update with each move. Assume all rewards are -1 except for the actions leading to the goal

state with respect to X-player.
Answer:
v(t)=v(t)+eta*(r+gama*v(t+1)-v(t))
The updated values are:
-0.286628
-0.280127
-0.264184
-0.570085
0.000000