

Question 1

Having the following XO sequence of states along with their values

	0.500
	0.537
	0.577
	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 along with their values

	0.500
	0.543
	0.506

	0.570
	0.509
	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:

$$v(t) = v(t) + \eta * (r + \gamma * v(t+1) - v(t))$$

The updated values are:

-0.011099

-0.005908

0.003613

0.008885

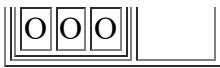
0.694618

1.000000

Question 3

Having the following XO sequence of states along with their values

	0.500
	0.598
	0.515
	0.576
	0.000



Assume a learning rate of 0.84 what will be updated values adopting gradient-based state value update with each move.

Answer:

$$v(t) = v(t) + \eta * (v(t+1) - v(t))$$

The updated values are:

0.582558

0.528692

0.566151

0.092130

0.000000

Question 4

Having the following XO sequence of states along with their values

	0.500
	0.567
	0.542
	1.000

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:

$$v(t) = v(t) + \eta * (r + \gamma * v(t+1) - v(t))$$

The updated values are:

-0.168895

-0.165704

0.834571

1.000000

Question 5

Having the following XO sequence of states along with their values

	0.500
	0.509
	0.507
	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 along with their values

	0.500
	0.549
	0.510
	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:

$$v(t) = v(t) + \eta * (r + \gamma * v(t+1) - v(t))$$

The updated values are:

-0.263209

-0.281383

-0.297320

0.500000

Question 7

Having the following XO sequence of states along with their values

	0.500
	0.543
	0.527
	1.000

Assume a learning rate of 0.87 what will be updated values adopting gradient-based state value update with each move.

Answer:

$$v(t) = v(t) + \eta * (v(t+1) - v(t))$$

The updated values are:

0.537694

0.529043

0.938498

1.000000

Question 8

Having the following XO sequence of states along with their values

	0.500
--	-------

	0.521
	0.541
	0.581
	0.542
	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:

$$v(t) = v(t) + \eta * (r + \gamma * v(t+1) - v(t))$$

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 along with their values

	0.500
	0.528
	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:

$$v(t) = v(t) + \eta * (v(t+1) - v(t))$$

The updated values are:

0.519947

0.517146

0.853684

1.000000

Question 10

Having the following XO sequence of states along with their values

O	O	-
X	X	-
-	X	O
O	O	X
X	X	-
-	X	O
O	O	X
X	X	O
-	X	O
O	O	X
X	X	O
X	X	O

0.500

0.510

0.552

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:

$$v(t) = v(t) + \eta * (r + \gamma * v(t+1) - v(t))$$

The updated values are:

-0.090582

-0.064632

0.755931

1.000000

Question 11

Having the following XO sequence of states along with their values

	0.500
	0.558
	0.511
	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 along with their values

	0.500
	0.519
	0.555
	0.500

O	X	O
-	-	X

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:

$$v(t) = v(t) + \eta * (r + \gamma * v(t+1) - v(t))$$

The updated values are:

-0.326055

-0.298000

-0.326432

0.500000

Question 13

Having the following XO sequence of states along 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:

$$v(t) = v(t) + \eta * (v(t+1) - v(t))$$

The updated values are:

0.538635

0.572285

0.949139

1.000000

Question 14

Having the following XO sequence of states along with their values

X - X	0.500
-----------	-------

	0.597
	0.558
	0.565
	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:

$$v(t) = v(t) + \eta * (r + \gamma * v(t+1) - v(t))$$

The updated values are:

-0.298166

-0.302966

-0.306758

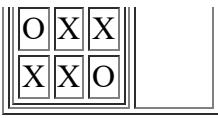
-0.671349

0.000000

Question 15

Having the following XO sequence of states along with their values

	0.500
	0.556
	0.586
	0.500



Assume a learning rate of 0.66 what will be updated values adopting gradient-based state value update with each move.

Answer:

$$v(t) = v(t) + \eta * (v(t+1) - v(t))$$

The updated values are:

0.536733

0.575601

0.529198

0.500000

Question 16

Having the following XO sequence of states along with their values

	0.500
	0.585
	0.526
	0.558
	0.558
	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:

$$v(t) = v(t) + \eta * (r + \gamma * v(t+1) - v(t))$$

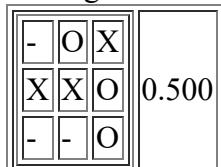
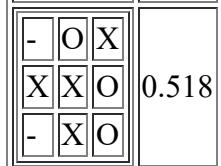
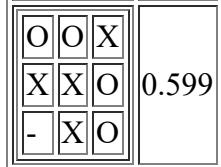
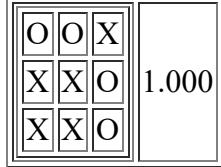
The updated values are:

-0.096810

-0.103134
 -0.104107
 -0.092256
 0.821847
 1.000000

Question 17

Having the following XO sequence of states along with their values

	0.500
	0.518
	0.599
	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

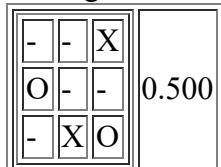
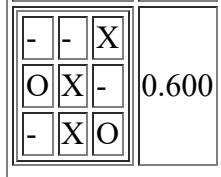
0.593752

0.971960

1.000000

Question 18

Having the following XO sequence of states along with their values

	0.500
	0.600

	0.596
	0.530
	0.590
	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:

$$v(t) = v(t) + \eta * (r + \gamma * v(t+1) - v(t))$$

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 along with their values

	0.500
	0.528
	0.533
	0.540

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

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 along with their values

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

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:

$$v(t) = v(t) + \eta * (r + \gamma * v(t+1) - v(t))$$

The updated values are:

-0.286628

-0.280127

-0.264184

-0.570085

0.000000