

Epsilon:

```
current_eps = max(0.01, epsilon - (0.01 * infoQ[state[0],  
state[1], state[2]]))
```

Learning Process started. Will finish when Episode = 15000

-----

-----

Policy with the weights: [1.0, 0.31, 0.013]

Expected value is: [7.0, 0.0, -3.375]

-----

The Learnt Policy has the following Value for alpha = 0.2 is:

[ 2.40454611 -1.14807813 -0.77043495]

-----

The Learnt Policy has the following Value for alpha = 0.25 is:

[ 4.05685348 -2.12280967 -3.08234753]

-----

The Learnt Policy has the following Value for alpha = 0.3 is:

[ 4.33572396 -1.07260332 -3.10005664]

-----

The Learnt Policy has the following Value for alpha = 0.35 is:

[ 3.36588977 -1.0003156 -2.70182737]

-----

The Learnt Policy has the following Value for alpha = 0.39999999999999997 is:

[ 3.97762721 -0.75883582 -4.06844502]

-----

The Learnt Policy has the following Value for alpha = 0.44999999999999996 is:

[ 4.90521759 -2.27435943 -4.27423456]

-----

The Learnt Policy has the following Value for alpha = 0.49999999999999994 is:

[ 4.35157192 -1.90133941 -2.73483407]

-----

The Learnt Policy has the following Value for alpha = 0.5499999999999999 is:

[ 5.135118 -2.2931228 -2.88300783]

-----

The Learnt Policy has the following Value for alpha = 0.5999999999999999 is:

[ 5.92575336 -4.91748909 -1.52807032]

-----

The Learnt Policy has the following Value for alpha = 0.6499999999999999 is:

[ 4.74268469 -4.3274921 -0.54920971]

-----

The Learnt Policy has the following Value for alpha = 0.7 is:

[ 4.48745516 -2.44482114 -1.28082665]

-----

The Learnt Policy has the following Value for alpha = 0.7499999999999998 is:

[ 4.87270429 -2.46780871 -0.98006883]

-----

The Learnt Policy has the following Value for alpha = 0.7999999999999998 is:

[ 7.74527987 -8.11235252 -3.39531822]

-----

-----

Policy with the weights: [1.0, 0.372, 1.5]

Expected value is: [6.03125, 0.0, 0.0]

-----

The Learnt Policy has the following Value for alpha = 0.2 is:

[ 2.91786472e+000 -7.48192077e-134 0.00000000e+000]

-----

The Learnt Policy has the following Value for alpha = 0.25 is:

[ 2.81918908 -1.33029263 0. ]

```

-----
The Learnt Policy has the following Value for alpha = 0.3 is:
[ 3.14401327 -2.07452401 0. ]
-----
The Learnt Policy has the following Value for alpha = 0.35 is:
[ 2.66355714 -2.83758089 0. ]
-----
The Learnt Policy has the following Value for alpha = 0.39999999999999997 is:
[ 4.50134680e+00 -3.18084236e+00 -7.87855933e-45]
-----
The Learnt Policy has the following Value for alpha = 0.44999999999999996 is:
[ 3.05581113e+00 -1.71378220e+00 -2.59389602e-09]
-----
The Learnt Policy has the following Value for alpha = 0.49999999999999994 is:
[ 3.60711103e+000 -1.62219975e+000 -9.23704378e-135]
-----
The Learnt Policy has the following Value for alpha = 0.5499999999999999 is:
[ 3.79056723e+00 -1.32021006e+00 -5.86385937e-32]
-----
The Learnt Policy has the following Value for alpha = 0.5999999999999999 is:
[ 4.76707426e+00 -1.76553955e+00 -9.70276133e-15]
-----
The Learnt Policy has the following Value for alpha = 0.6499999999999999 is:
[ 4.22138777 -2.37949 0. ]
-----
The Learnt Policy has the following Value for alpha = 0.7 is:
[ 4.93498160e+000 -1.09598544e+000 -3.19042266e-258]
-----
The Learnt Policy has the following Value for alpha = 0.7499999999999998 is:
[ 4.19067209e+00 -2.01834030e+00 -1.27890961e-61]
-----
The Learnt Policy has the following Value for alpha = 0.7999999999999998 is:
[ 4.82489578e+00 -5.62423587e-01 -2.35816899e-62]
-----

-----
Policy with the weights: [1.0, 0.09, 1e-09]
Expected value is: [10.0, -20.0, -4.0]
-----
The Learnt Policy has the following Value for alpha = 0.2 is:
[ 8.29395276 -15.10772095 -1.80572453]
-----
The Learnt Policy has the following Value for alpha = 0.25 is:
[ 7.28865003 -14.90987218 -1.59467708]
-----
The Learnt Policy has the following Value for alpha = 0.3 is:
[ 9.39536522 -19.98119786 -5.26621697]
-----
The Learnt Policy has the following Value for alpha = 0.35 is:
[ 8.27214792 -20.53206747 -2.98623797]
-----
The Learnt Policy has the following Value for alpha = 0.39999999999999997 is:
[ 8.19511902 -18.11844964 -3.53492808]
-----
The Learnt Policy has the following Value for alpha = 0.44999999999999996 is:
[ 9.15060382 -17.92216633 -3.16114659]
-----
The Learnt Policy has the following Value for alpha = 0.49999999999999994 is:
[ 6.78971364 -16.10602947 -1.98568227]
-----

```

The Learnt Policy has the following Value for alpha = 0.5499999999999999 is:  
[ 7.5245923 -11.29752049 -4.87040642]  
-----

The Learnt Policy has the following Value for alpha = 0.5999999999999999 is:  
[ 7.72660423 -14.66178267 -3.23703378]  
-----

The Learnt Policy has the following Value for alpha = 0.6499999999999999 is:  
[ 8.10230001 -14.84101836 -2.70514582]  
-----

The Learnt Policy has the following Value for alpha = 0.7 is:  
[ 9.16444594 -24.58611849 -1.43786937]  
-----

The Learnt Policy has the following Value for alpha = 0.7499999999999998 is:  
[ 7.5623236 -13.53250781 -2.64773983]  
-----

The Learnt Policy has the following Value for alpha = 0.7999999999999998 is:  
[ 8.25634674 -15.3699815 -3.61868085]  
-----

-----  
Policy with the weights: [1.0, 0.045, 0.233]  
Expected value is: [10.0, -30.0, -0.75]  
-----

The Learnt Policy has the following Value for alpha = 0.2 is:  
[ 7.61920388e+000 -1.84183605e+001 -1.97626258e-323]  
-----

The Learnt Policy has the following Value for alpha = 0.25 is:  
[ 8.7538904 -19.97348759 -1.10800286]  
-----

The Learnt Policy has the following Value for alpha = 0.3 is:  
[ 9.74178921 -25.23956533 -1.3312465 ]  
-----

The Learnt Policy has the following Value for alpha = 0.35 is:  
[ 7.87571548 -15.94980818 -0.40257975]  
-----

The Learnt Policy has the following Value for alpha = 0.39999999999999997 is:  
[ 8.75910565 -15.19850486 -0.40625749]  
-----

The Learnt Policy has the following Value for alpha = 0.44999999999999996 is:  
[ 7.91150852 -10. -0.2817553 ]  
-----

The Learnt Policy has the following Value for alpha = 0.49999999999999994 is:  
[ 7.51735899e+000 -1.56616462e+001 -9.88131292e-324]  
-----

The Learnt Policy has the following Value for alpha = 0.5499999999999999 is:  
[ 6.98045633 -10. -0.20149107]  
-----

The Learnt Policy has the following Value for alpha = 0.5999999999999999 is:  
[ 8.61634377 -20.66438703 -0.18418713]  
-----

The Learnt Policy has the following Value for alpha = 0.6499999999999999 is:  
[ 7.85922856e+000 -1.00000000e+001 -3.34040258e-132]  
-----

The Learnt Policy has the following Value for alpha = 0.7 is:  
[ 8.3536203 -21.50210167 -0.1058581 ]  
-----

The Learnt Policy has the following Value for alpha = 0.7499999999999998 is:  
[ 7.63445634 -10. -0.66815062]  
-----

The Learnt Policy has the following Value for alpha = 0.7999999999999998 is:

[ 8.98003629 -15.1581248 -0.31817689]

-----

-----

Policy with the weights: [1.0, 0.372, 1.5]

Expected value is: [6.03125, 0.0, 0.0]

-----

The Learnt Policy has the following Value for alpha = 0.2 is:

[ 2.64449029 -0.06821389 0. ]

-----

The Learnt Policy has the following Value for alpha = 0.25 is:

[ 2.73218549e+00 -7.06965049e-03 -5.58522367e-21]

-----

The Learnt Policy has the following Value for alpha = 0.3 is:

[ 2.55193859e+00 -2.39520622e+00 -1.21326602e-25]

-----

The Learnt Policy has the following Value for alpha = 0.35 is:

[ 2.42756201 -2.05869651 0. ]

-----

The Learnt Policy has the following Value for alpha = 0.39999999999999997 is:

[ 3.37876048 -1.96553415 0. ]

-----

The Learnt Policy has the following Value for alpha = 0.44999999999999996 is:

[ 4.28932295e+00 -2.01855642e+00 -2.39168886e-84]

-----

The Learnt Policy has the following Value for alpha = 0.49999999999999994 is:

[ 3.72623033e+00 -1.71767263e+00 -7.92373544e-26]

-----

The Learnt Policy has the following Value for alpha = 0.5499999999999999 is:

[ 4.23395227e+000 -1.96150839e+000 -3.53697617e-278]

-----

The Learnt Policy has the following Value for alpha = 0.5999999999999999 is:

[ 4.47192586e+000 -1.42076059e+000 -3.95252517e-323]

-----

The Learnt Policy has the following Value for alpha = 0.6499999999999999 is:

[ 4.09731956e+000 -2.61824923e-001 -5.30796361e-116]

-----

The Learnt Policy has the following Value for alpha = 0.7 is:

[ 4.81844871e+00 -2.24920079e+00 -6.64398726e-40]

-----

The Learnt Policy has the following Value for alpha = 0.7499999999999998 is:

[ 3.92042114e+00 -1.00181536e+00 -1.46716791e-45]

-----

The Learnt Policy has the following Value for alpha = 0.7999999999999998 is:

[ 4.98183338e+000 -1.54271773e+000 -1.75427754e-210]

-----

-----

Policy with the weights: [1.0, 0.005, 0.633]

Expected value is: [9.625, -30.0, 0.0]

-----

The Learnt Policy has the following Value for alpha = 0.2 is:

[ 8.89332468e+000 -2.34568717e+001 -1.97626258e-323]

-----

The Learnt Policy has the following Value for alpha = 0.25 is:

[ 9.71851224 -30. 0. ]

This alpha works!

-----

The Learnt Policy has the following Value for alpha = 0.3 is:

[ 8.72945189e+000 -2.37551828e+001 -9.88131292e-324]

-----  
The Learnt Policy has the following Value for alpha = 0.35 is:  
[ 8.67792406 -21.72473864 -0.62925231]  
-----

The Learnt Policy has the following Value for alpha = 0.39999999999999997 is:  
[ 6.66815781 -12.43851749 0. ]  
-----

The Learnt Policy has the following Value for alpha = 0.44999999999999996 is:  
[ 8.62110795e+000 -2.74222893e+001 -1.97626258e-323]  
-----

The Learnt Policy has the following Value for alpha = 0.49999999999999994 is:  
[ 8.01561357e+000 -2.06333832e+001 -1.97626258e-323]  
-----

The Learnt Policy has the following Value for alpha = 0.5499999999999999 is:  
[ 7.57739260e+00 -2.09560805e+01 -1.32496452e-94]  
-----

The Learnt Policy has the following Value for alpha = 0.5999999999999999 is:  
[ 8.57070472 -21.31903799 -0.31746027]  
-----

The Learnt Policy has the following Value for alpha = 0.6499999999999999 is:  
[ 8.97281824 -25.59160714 0. ]  
-----

The Learnt Policy has the following Value for alpha = 0.7 is:  
[ 7.93582277e+00 -1.86744378e+01 -1.25635717e-82]  
-----

The Learnt Policy has the following Value for alpha = 0.7499999999999998 is:  
[ 7.58368533e+000 -1.73462241e+001 -1.77068352e-111]  
-----

The Learnt Policy has the following Value for alpha = 0.7999999999999998 is:  
[ 7.53528648e+000 -1.94790143e+001 -1.97626258e-323]  
-----

Finnished!!!  
[[, ], [0.25]]

Process finished with exit code 0