

```

if episode > 10000:
    alpha = 0.2
elif episode > 20000:
    alpha = 0.1
elif episode > 30000:
    alpha = 0.05

```

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:
[6.69309106 -5.08768653 -5.2747319]

The Learnt Policy has the following Value for alpha = 0.25 is:
[6.40666681 -3.28440498 -4.68339419]

The Learnt Policy has the following Value for alpha = 0.3 is:
[5.72951476 -2.43298628 -4.09128589]

The Learnt Policy has the following Value for alpha = 0.35 is:
[5.8859642 -2.64083002 -2.11929936]

The Learnt Policy has the following Value for alpha = 0.39999999999999997 is:
[6.15507886 -2.16103711 -3.53386427]

The Learnt Policy has the following Value for alpha = 0.44999999999999996 is:
[7.36145638 -3.35633122 -2.45469386]

The Learnt Policy has the following Value for alpha = 0.49999999999999994 is:
[7.69808127 -4.71815969 -2.16306793]

The Learnt Policy has the following Value for alpha = 0.5499999999999999 is:
[6.53851825 -2.8657499 -5.02697282]

The Learnt Policy has the following Value for alpha = 0.5999999999999999 is:
[7.62226614 -3.95303139 -3.49971563]

The Learnt Policy has the following Value for alpha = 0.6499999999999999 is:
[7.20584427 -1.1797323 -3.51423125]

The Learnt Policy has the following Value for alpha = 0.7 is:
[7.31918911 -2.55039676 -3.28764066]

The Learnt Policy has the following Value for alpha = 0.7499999999999998 is:
[6.98716570e+00 -8.24601369e-45 -4.05063155e+00]

The Learnt Policy has the following Value for alpha = 0.7999999999999998 is:
[6.99773292 -0.00944473 -4.45638707]

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:
[ 4.72150374e+00 -1.84930791e+00 -1.56879467e-16]
-----
The Learnt Policy has the following Value for alpha = 0.25 is:
[ 5.61176273 -3.11423239 0. ]
-----
The Learnt Policy has the following Value for alpha = 0.3 is:
[ 4.41418908e+00 -8.76451592e-01 -5.23696219e-15]
-----
The Learnt Policy has the following Value for alpha = 0.35 is:
[ 4.45313121e+00 -4.52760680e-01 -2.06269110e-06]
-----
The Learnt Policy has the following Value for alpha = 0.39999999999999997 is:
[ 5.58810808e+00 -2.02782349e+00 -3.01127545e-22]
-----
The Learnt Policy has the following Value for alpha = 0.44999999999999996 is:
[ 5.95261667e+00 -1.55449611e+00 -1.90306706e-26]
-----
The Learnt Policy has the following Value for alpha = 0.49999999999999994 is:
[ 6.13363753e+00 -8.99042543e-01 -2.08079435e-15]
-----
The Learnt Policy has the following Value for alpha = 0.54999999999999999 is:
[ 5.34758548e+00 -1.33287535e+00 -1.32434972e-06]
-----
The Learnt Policy has the following Value for alpha = 0.59999999999999999 is:
[ 5.74068887e+00 -1.14423839e+00 -2.61203104e-12]
-----
The Learnt Policy has the following Value for alpha = 0.64999999999999999 is:
[ 6.08394423e+00 -8.81808972e-01 -6.41329052e-10]
-----
The Learnt Policy has the following Value for alpha = 0.7 is:
[ 6.22810255e+00 -1.10712264e+00 -4.33170075e-30]
-----
The Learnt Policy has the following Value for alpha = 0.74999999999999998 is:
[ 5.26472210e+00 -3.61481161e-08 -6.13008379e-09]
-----
The Learnt Policy has the following Value for alpha = 0.79999999999999998 is:
[ 6.09076762e+00 -9.16040793e-18 -4.81475014e-06]
This alpha works!
-----

```

```

-----
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:
[ 8.26999717 -12.08037937 -6.01265751]
-----
The Learnt Policy has the following Value for alpha = 0.25 is:
[ 9.48952702 -17.94913963 -2.29296575]
-----
The Learnt Policy has the following Value for alpha = 0.3 is:
[ 8.49552218 -10. -2.09433467]
-----
The Learnt Policy has the following Value for alpha = 0.35 is:
[ 8.81018332 -21.70245751 -5.29367242]
-----
The Learnt Policy has the following Value for alpha = 0.39999999999999997 is:
[ 7.97857888 -17.36633269 -5.48845445]

```

The Learnt Policy has the following Value for alpha = 0.44999999999999996 is:
[8.99994864 -17.93863563 -3.83890852]

The Learnt Policy has the following Value for alpha = 0.49999999999999994 is:
[10. -21.60386871 -2.83001804]

The Learnt Policy has the following Value for alpha = 0.5499999999999999 is:
[9.79217843 -17.93370669 -3.48030038]

The Learnt Policy has the following Value for alpha = 0.5999999999999999 is:
[8.99706282 -13.81176781 -6.95635981]

The Learnt Policy has the following Value for alpha = 0.6499999999999999 is:
[9. -10. -3.62301797]

The Learnt Policy has the following Value for alpha = 0.7 is:
[10. -20. -4.03983987]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.7499999999999998 is:
[9. -10. -1.34079211]

The Learnt Policy has the following Value for alpha = 0.7999999999999998 is:
[9. -10. -1.38357531]

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:
[10. -30. -0.29432669]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.25 is:
[8.68147934 -10. -0.56178503]

The Learnt Policy has the following Value for alpha = 0.3 is:
[9.82507091 -29.99495316 -0.23010431]

The Learnt Policy has the following Value for alpha = 0.35 is:
[10. -30. -0.4308863]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.39999999999999997 is:
[8.45625794 -10. -0.21927124]

The Learnt Policy has the following Value for alpha = 0.44999999999999996 is:
[9.72567741 -28.59907903 -1.3772671]

The Learnt Policy has the following Value for alpha = 0.49999999999999994 is:
[8.45442994 -11.89803584 -1.30441403]

The Learnt Policy has the following Value for alpha = 0.5499999999999999 is:
[9.01102929 -10.22294978 -0.92746194]

The Learnt Policy has the following Value for alpha = 0.5999999999999999 is:
[9.65399449 -23.07988982 -1.03801653]

```

The Learnt Policy has the following Value for alpha = 0.6499999999999999 is:
[ 9.      -10.      -0.75822405]
-----
The Learnt Policy has the following Value for alpha = 0.7 is:
[ 9.      -10.      -0.47108435]
-----
The Learnt Policy has the following Value for alpha = 0.7499999999999998 is:
[ 10.     -30.     -0.65343398]
This alpha works!
-----
The Learnt Policy has the following Value for alpha = 0.7999999999999998 is:
[ 9.      -10.      -0.70322189]
-----

-----
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:
[ 4.51197071e+000 -1.07642422e+000 -2.27683353e-201]
-----
The Learnt Policy has the following Value for alpha = 0.25 is:
[ 4.91750684 -0.93790381  0.      ]
-----
The Learnt Policy has the following Value for alpha = 0.3 is:
[ 4.33945762e+00 -8.47698313e-01 -1.46753901e-30]
-----
The Learnt Policy has the following Value for alpha = 0.35 is:
[ 4.98919390e+000 -1.34942584e+000 -9.88131292e-324]
-----
The Learnt Policy has the following Value for alpha = 0.39999999999999997 is:
[ 6.29872790e+00 -2.70805103e+00 -2.05047185e-27]
-----
The Learnt Policy has the following Value for alpha = 0.44999999999999996 is:
[ 5.82364196e+00 -1.13198725e+00 -3.86243472e-20]
-----
The Learnt Policy has the following Value for alpha = 0.49999999999999994 is:
[ 5.82329889e+00 -2.20826610e+00 -8.69301419e-13]
-----
The Learnt Policy has the following Value for alpha = 0.5499999999999999 is:
[ 6.82950261e+00 -3.36575066e+00 -1.89730576e-70]
-----
The Learnt Policy has the following Value for alpha = 0.5999999999999999 is:
[ 6.02590400e+00 -1.05285182e+00 -1.17824619e-38]
-----
The Learnt Policy has the following Value for alpha = 0.6499999999999999 is:
[ 6.01506379e+00 -3.03605059e-09 -3.33977373e-19]
This alpha works!
-----
The Learnt Policy has the following Value for alpha = 0.7 is:
[ 5.03592331e+00 -2.14884917e-01 -3.43330000e-07]
-----
The Learnt Policy has the following Value for alpha = 0.7499999999999998 is:
[ 6.05518126e+00 -2.35734828e-01 -8.43851523e-08]
This alpha works!
-----
The Learnt Policy has the following Value for alpha = 0.7999999999999998 is:
[ 5.92675215e+00 -7.50876423e-28 -2.45127050e-05]
This alpha works!
-----

```

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:
[9.90352158e+00 -3.00000000e+01 -2.17003831e-45]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.25 is:
[9.31972788e+00 -3.00000000e+01 -1.05709034e-67]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.3 is:
[9.69877439e+000 -3.00000000e+001 -8.65037870e-120]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.35 is:
[9.48617531e+00 -3.00000000e+01 -1.15578287e-86]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.39999999999999997 is:
[9.34816533e+00 -3.00000000e+01 -3.61772414e-59]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.44999999999999996 is:
[8.95006957e+000 -1.60124073e+001 -1.13391751e-247]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.49999999999999994 is:
[9.72017468e+00 -3.00000000e+01 -5.43444113e-10]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.5499999999999999 is:
[9.60006944e+000 -3.00000000e+001 -6.53160516e-150]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.5999999999999999 is:
[9.70270320e+00 -3.00000000e+01 -6.68878851e-66]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.6499999999999999 is:
[9.49820526e+00 -2.68897863e+01 -6.66686951e-51]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.7 is:
[9.60058011e+00 -3.00000000e+01 -1.57799570e-73]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.7499999999999998 is:
[9.52594547e+00 -3.00000000e+01 -5.19559839e-44]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.7999999999999998 is:
[9.25881972e+00 -2.99999979e+01 -1.10812719e-18]
This alpha works!

Finnished!!!

[[, [0.8], [0.7], [0.2, 0.35, 0.75], [0.65, 0.75, 0.8], [0.2, 0.25, 0.3, 0.35, 0.4, 0.5, 0.55, 0.6, 0.7, 0.75, 0.8]]

Process finished with exit code 0