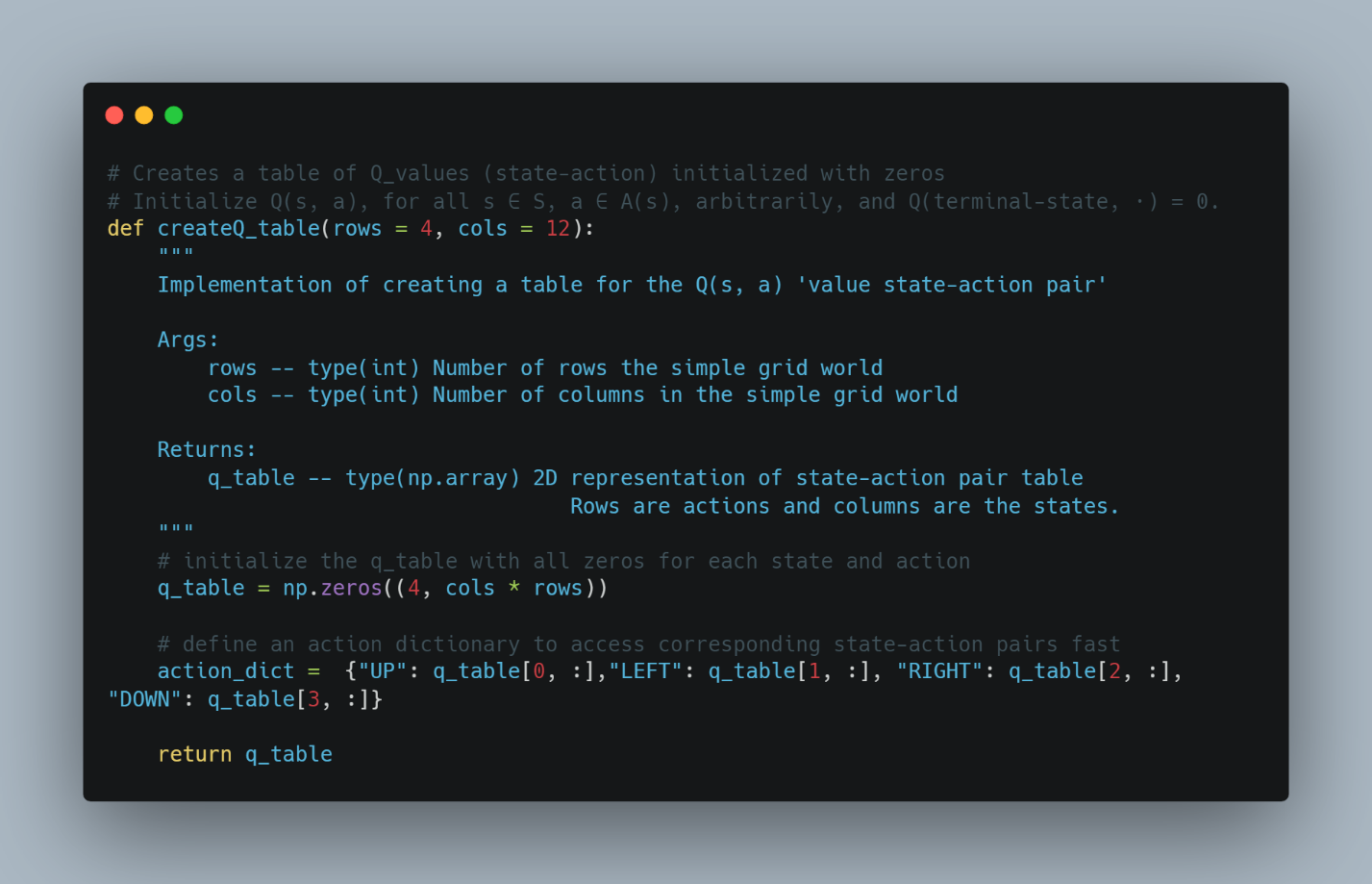
به نام خدا

پروژه پنجم درس هوش مصنوعي – RL بهاره كاوسي نژاد – 99431217

**Walking Cliff:**

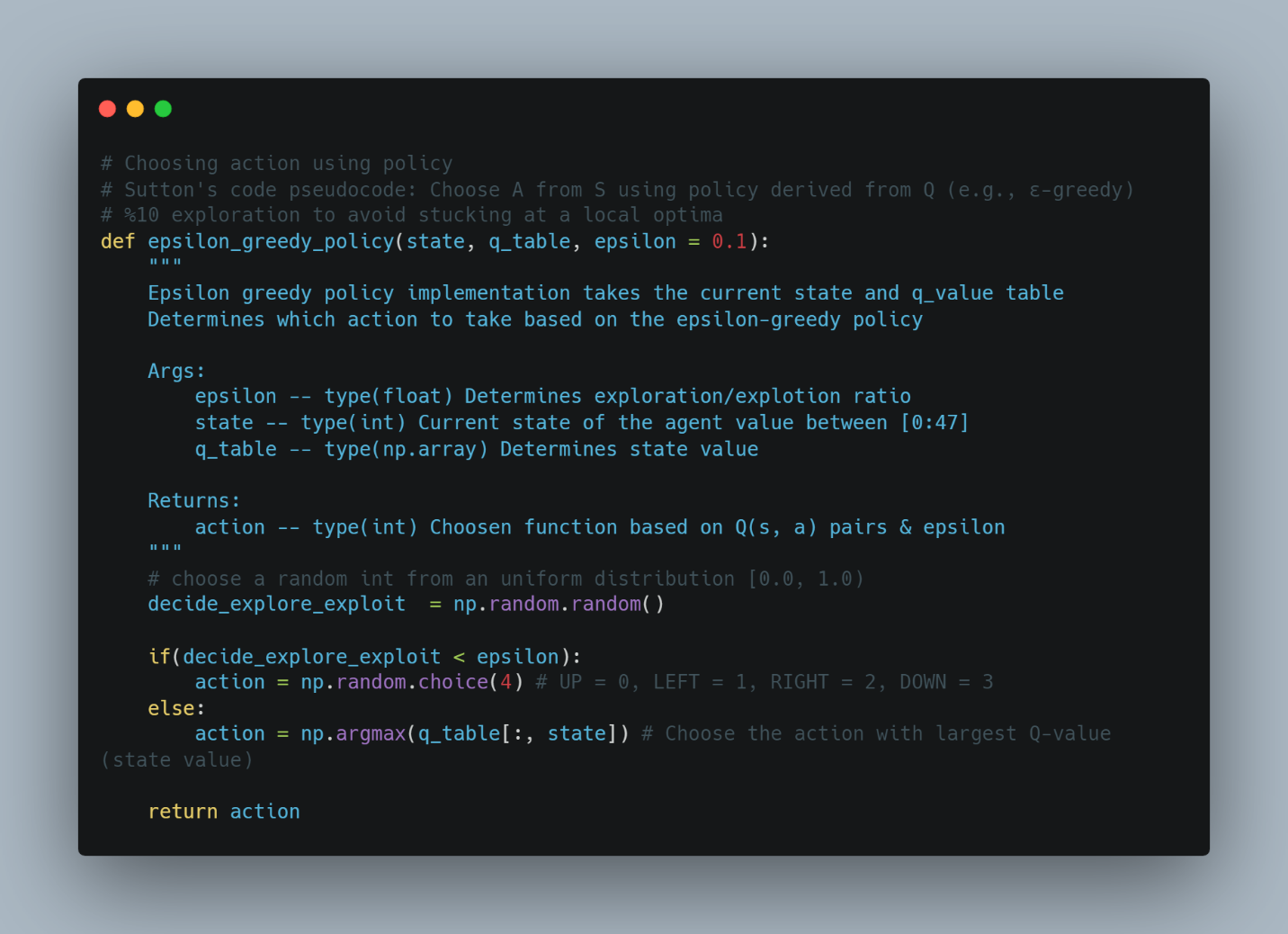


ابتدا كتابخانه هاي لازم را import مي‌كنيم.



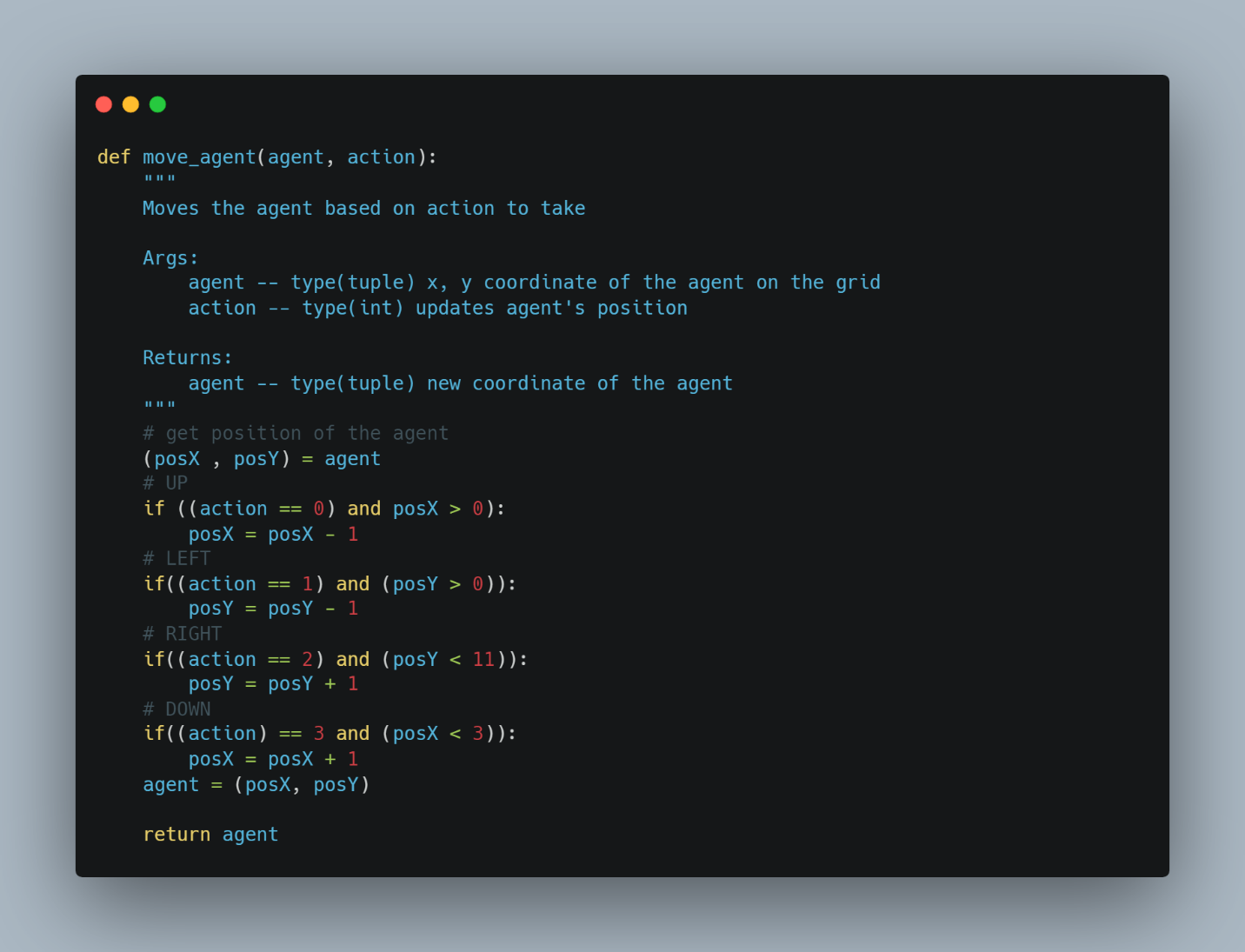
**ساختن جدول Q\_values:** مقادیر این جدول در ابتدا صفر هستند. Q(s, a) را براي تمامي stateها و actionها به صورت دلخواه مقدار دهي مي‌كنيم. و Q(terminal-state, 0) را صفر قرار مي‌دهيم. اين تابع به عنوان ورودي تعداد سطرها و ستون‌ها را دريافت مي‌كند و به عنوان خروجي q\_table را باز مي‌گرداند كه يك آرايه 2 بعدي است و سطرها نمايانگر actionها و ستون‌ها نمايانگر stateها هستند.

همچنين يك action dictionary براي يافتن سريع جفت‌هاي state-action تعريف مي‌كنيم.

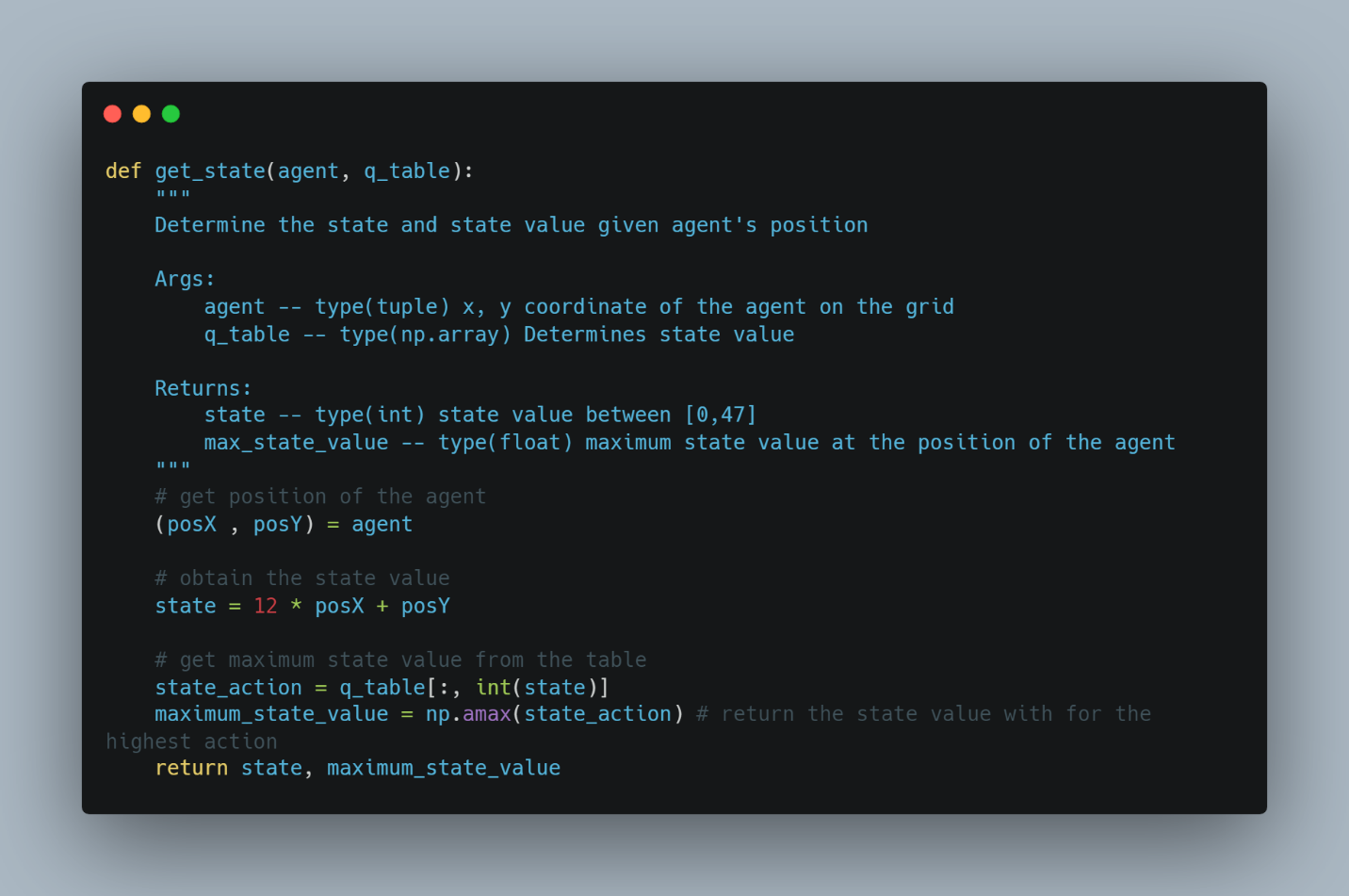


**انتخاب یک action بر اساس policy:** بر اساس policy يك action انتخاب مي‌شود. اين policy از روي Q مشخص مي‌شود. از 10 درصد exploration براي جلوگيري از گير كردن در يك local optima استفاده مي‌شود.

اين تابع state كنوني و جدول q\_value را مي‌گيرد و بر اساس policy تعيين مي كند كه چه actionاي را انجام دهد.



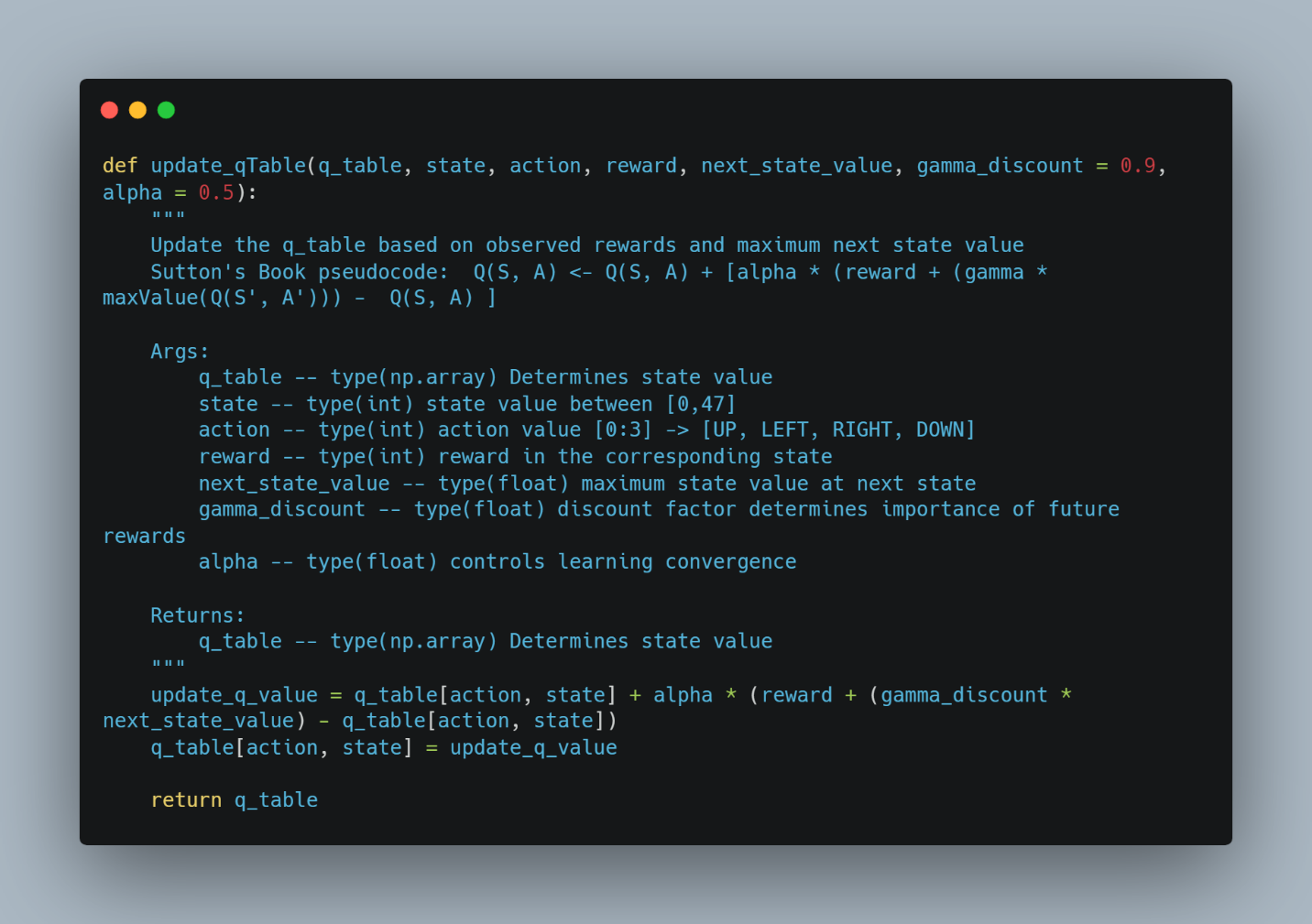
**حرکت دادن agent:** بر اساس action انتخاب شده agent را حركت مي‌دهد.



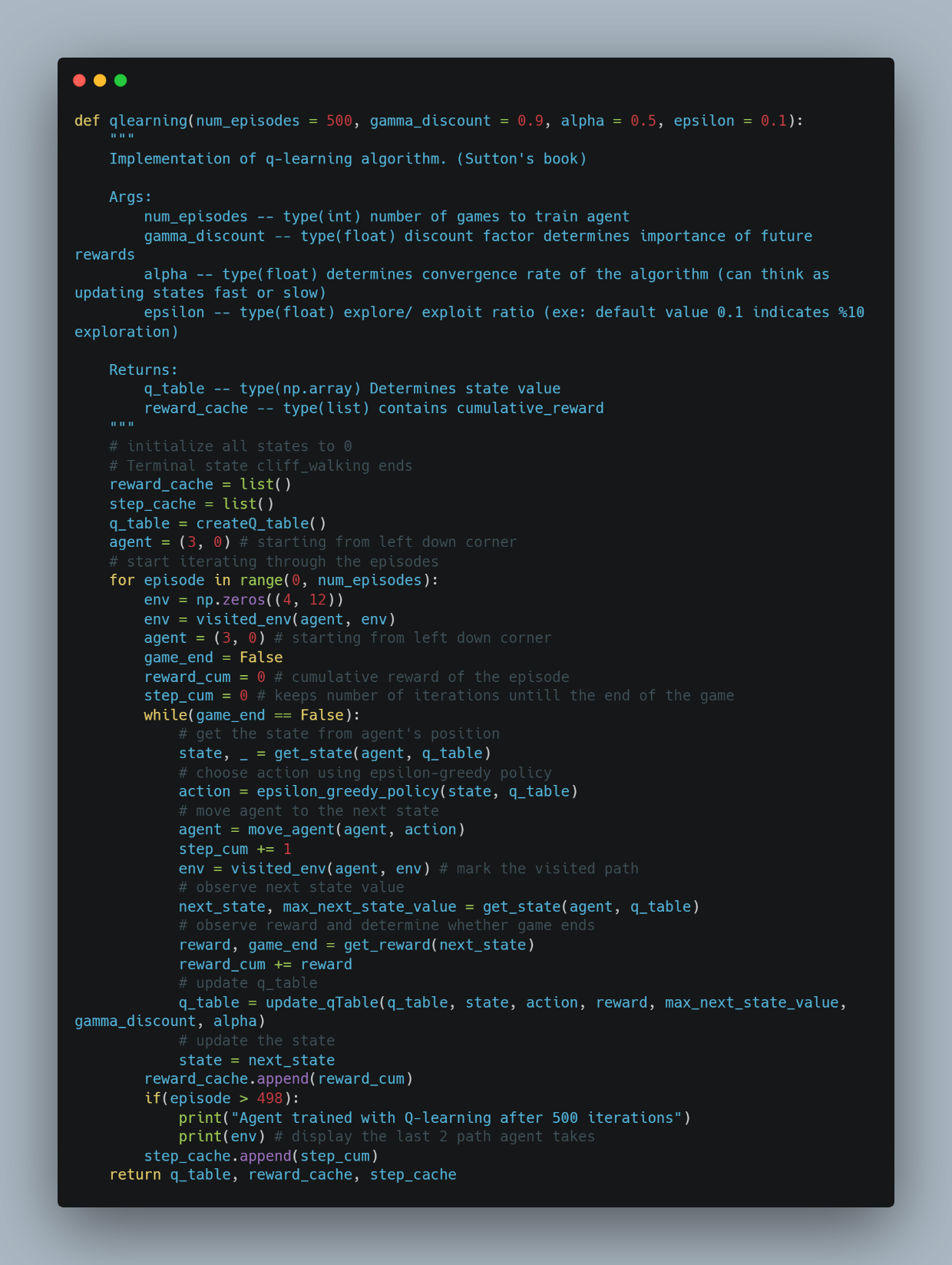
**پیدا کردن موقعیت state:** بر اساس موقعيت agent، state و state value را برمي‌گرداند.



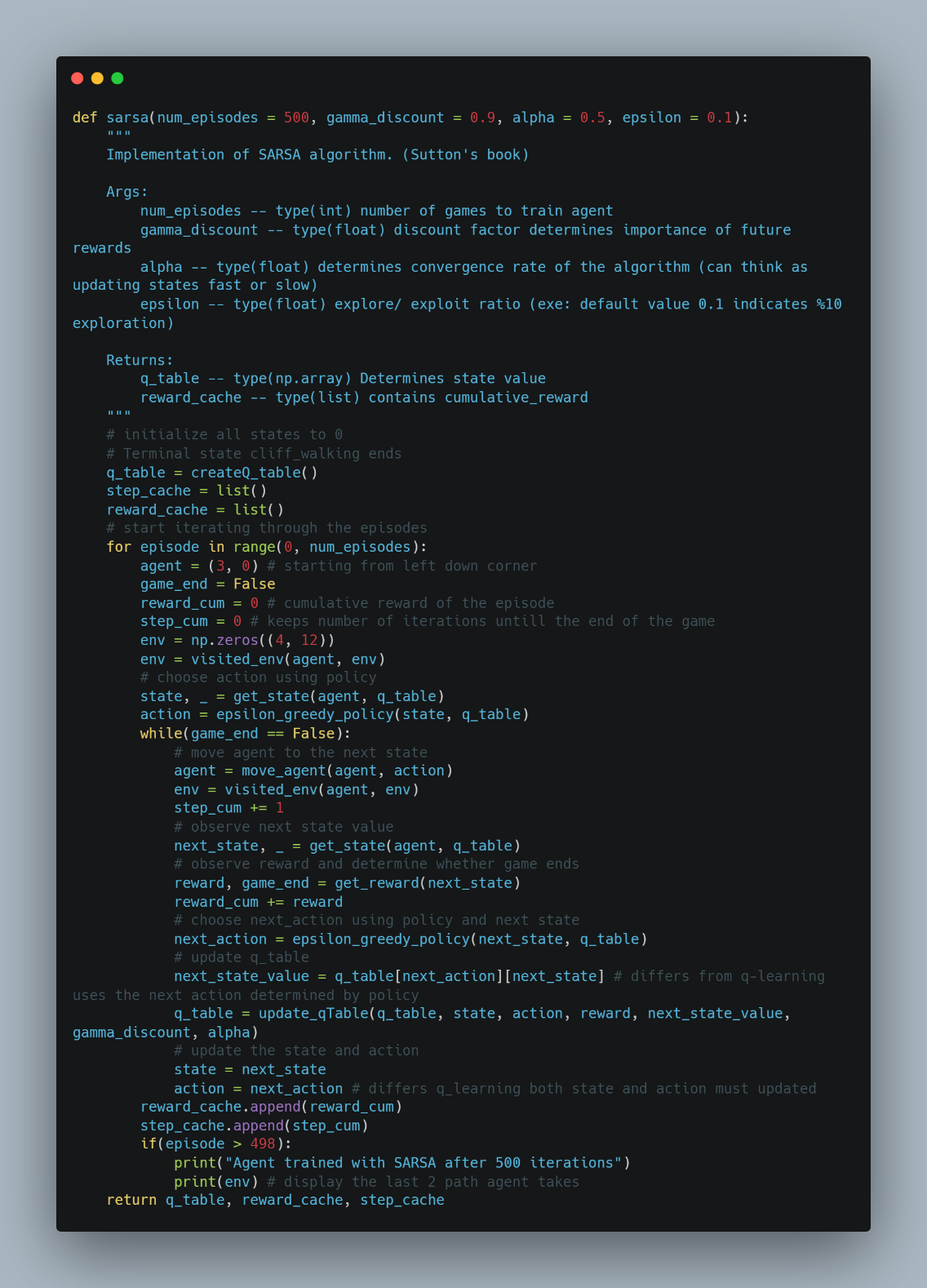
**دریافت reward:** اين تابع مقدار reward را در state ورودي باز مي‌گرداند. همچنين اين تابع با يك flag با عنوان game\_end پايان بازي را مشخص مي‌كند. همه stateها به جز cliffها مقدار -1 دارند. اگر agent با يك cliff برخورد كند جريمه مي‌شود و بازي تمام مي‌شود.



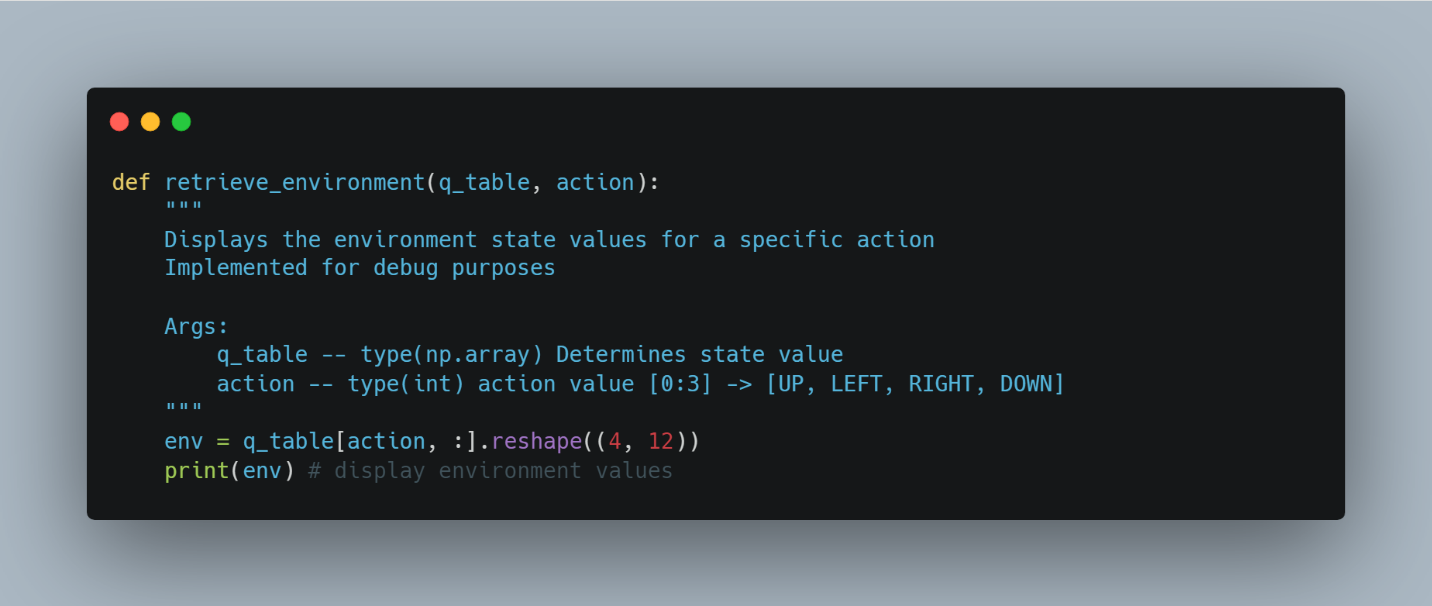
از این تابع برای آپدیت کردن Q Table استفاده می شود.



تابع اصلی Q Learning



تابع پیاده سازی sarsa



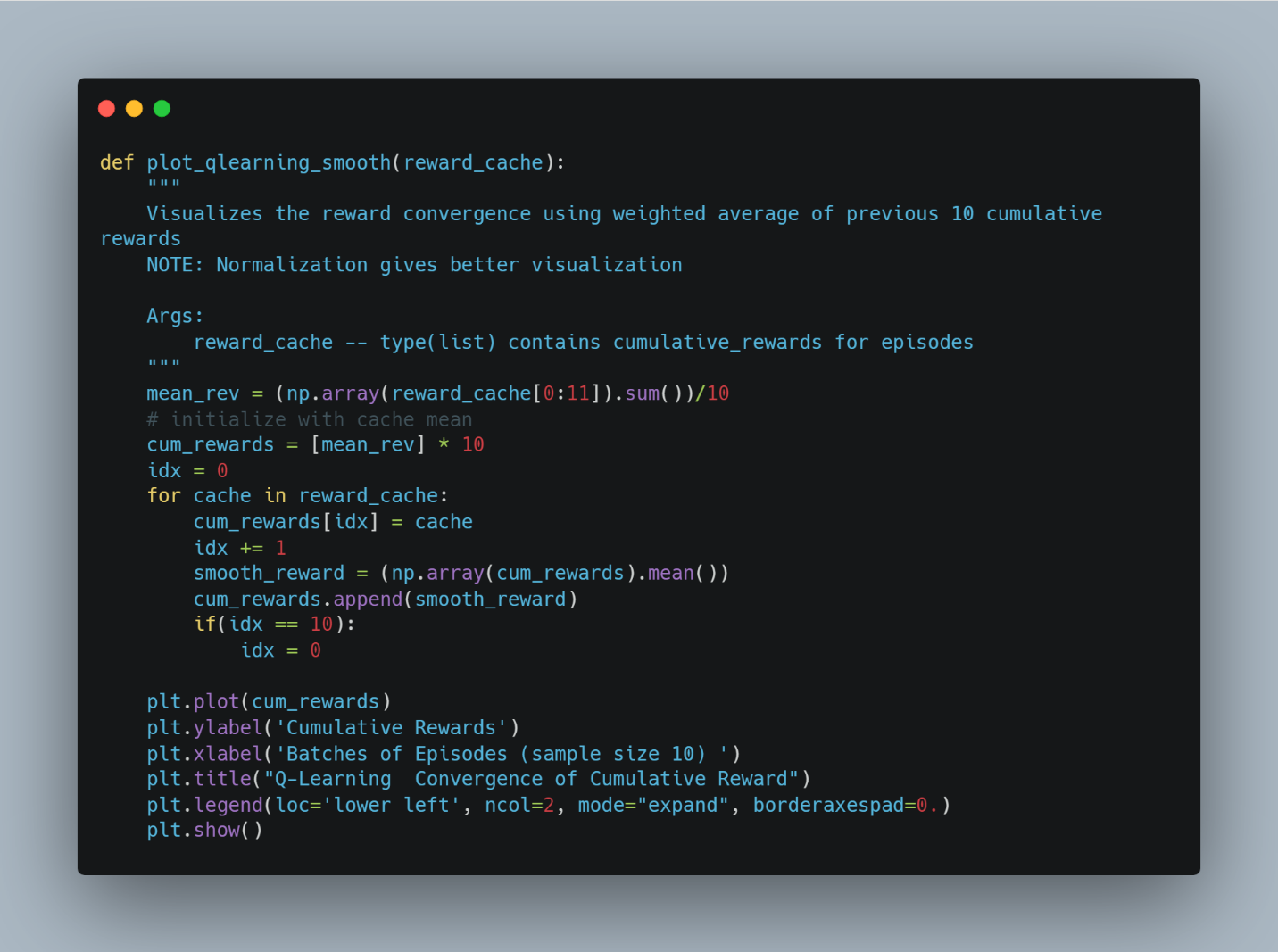
توابع مربوط به environment



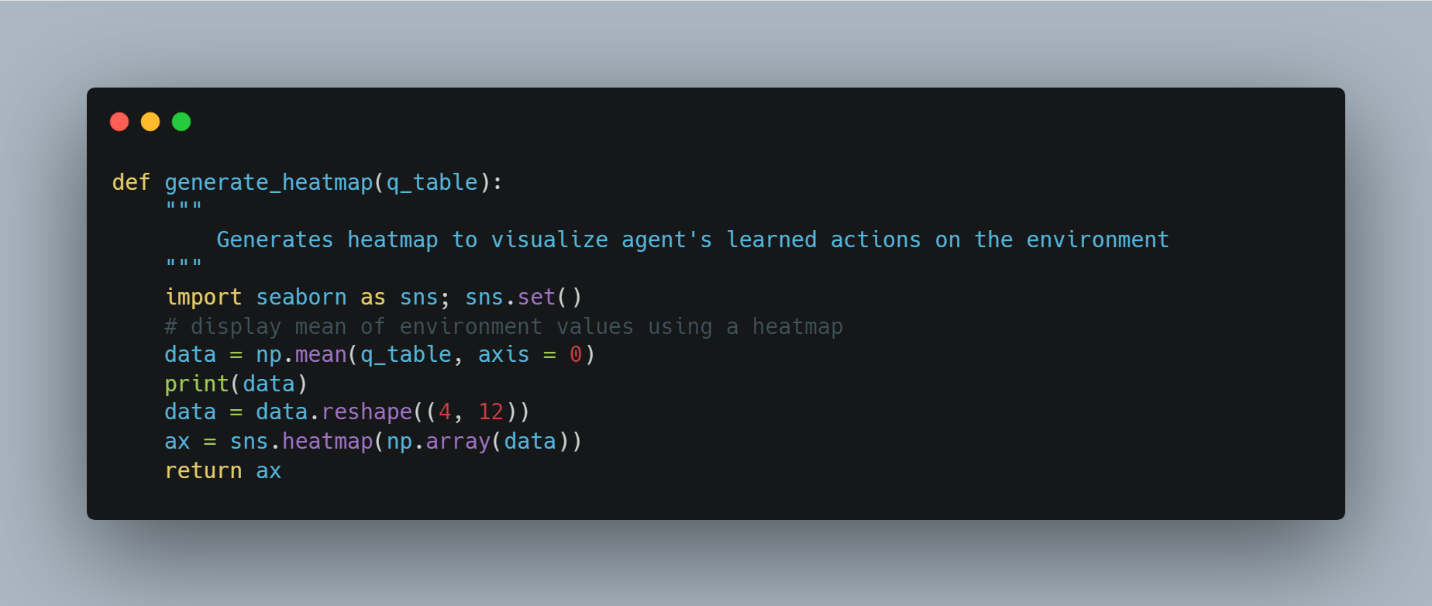
ترسیم



ترسیم نمودار تعداد stepها



ترسیم نمودار qLearning



ترسیم نمودار heatmap



تابع main

**نتایج:**

Agent trained with SARSA after 500 iterations

[[1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 0.]

[1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0.]

[1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1.]

[1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.]]

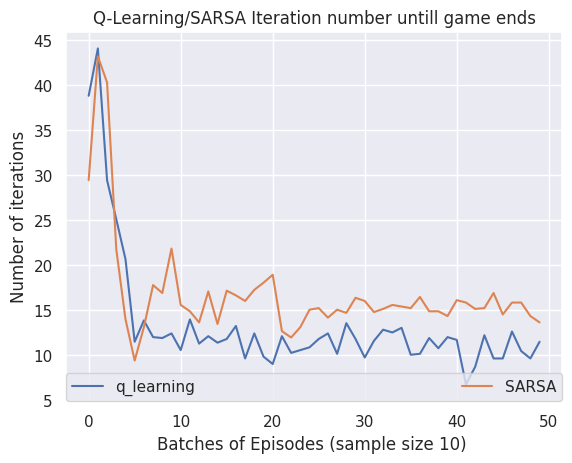
Agent trained with Q-learning after 500 iterations

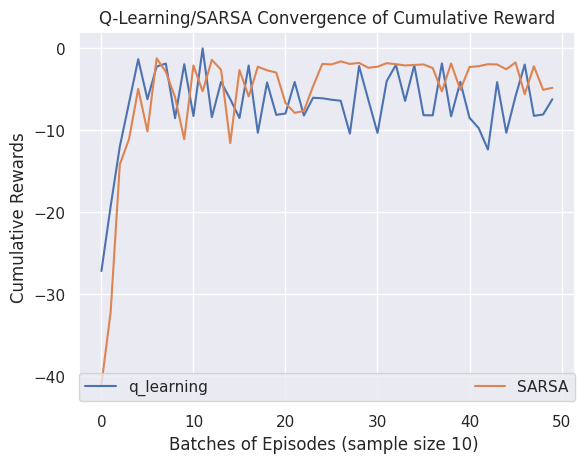
[[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

[1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.]

[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.]]





Visualize environment Q-learning

[-5.73053569e+00 -5.47502483e+00 -5.08353013e+00 -4.77449754e+00

-4.36219148e+00 -3.69903930e+00 -2.89779470e+00 -2.03160112e+00

-1.71358850e+00 -1.22648939e+00 -5.85761522e-01 -1.14718750e+00

-5.39601117e+00 -4.69657016e+00 -4.53954103e+00 -3.99155109e+00

-2.99937957e+00 -3.08770223e+00 -1.28004843e+00 -1.00044170e-01

6.45914989e-01 -2.87698007e-03 1.19852452e+00 6.80603435e-01

-4.47838404e+00 -2.79687128e+01 -2.74478423e+01 -2.68636111e+01

-2.62131708e+01 -2.55374207e+01 -2.47770465e+01 -2.38837981e+01

-2.21839393e+01 -2.19045331e+01 -2.07522152e+01 6.79399672e+00

-2.85465868e+01 0.00000000e+00 0.00000000e+00 0.00000000e+00

0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00

0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]

Axes(0.125,0.11;0.62x0.77)

Visualize SARSA

[ -6.69299486 -6.1686545 -5.86063132 -5.17941289 -4.64015444

-3.91514395 -3.0877163 -2.4865526 -1.29300736 0.18927052

1.62869468 2.9576164 -6.96140725 -8.41846674 -9.84603984

-5.35505833 -4.70887744 -4.14651896 -3.68223545 -7.71126553

-4.79582634 -3.70311474 2.19069024 4.82379737 -18.41668353

-34.02226785 -26.31921599 -25.12427517 -23.74818571 -21.95838849

-21.47388817 -17.78334869 -20.93004282 -21.53369002 -21.18271946

6.83491462 -41.48827017 0. 0. 0.

0. 0. 0. 0. 0.

0. 0. 0. ]

Axes(0.125,0.11;0.496x0.77)

