Search or jump to... Pull requests Issues Marketplace Explore

Learn Git and GitHub without any code!

Using the Hello World guide, you'll learn a branch, write commits, and open a pull request.

Start the guide

langgape / qq / ml_learn

master / ml_learn / homework_5 / q_table_q_learning_guide.ipynb Go to file

langgape 1 minute ago

Latest commit c1a4ddc 1 minute ago History

No LFS pointer found.

0001 lines 0000 lines 0001 - 00 00 Raw Blame

Q-Learning

In this homework, you should complete the Q-learning algorithm for the environment gym-taxi

In [1]:
```
import gym
import numpy as np
import matplotlib

# set the backend to inline in ipynb mode(%) or remove the objective
%matplotlib inline
```

Step 1 Creat the environment

Using the OpenAI Gym environment gym-taxi

In [2]:
```
env = gym.make('Taxi-v3')
env.reset()
```

Out[2]:
```
0
```

Step 2 Create the Q-table and initialize it

You can see that gym-taxi has the dimensions of state space and action space

In [3]:
```
# [observation_space_n, action_space_n]
print(env.observation_space.n)
print(env.action_space.n)

# Create required Q(s) for observations in this env
Q = np.zeros([env.observation_space.n, env.action_space.n])
```

```
500
6
```

Step 3 Configure the hyperparameters

In [4]:
```
total_episodes = 5
max_steps = 100
learning_rate = 0.7

gamma = 0.618

# Exploration parameters
epsilon = 1.0
max_epsilon = 1.0
min_epsilon = 0.01
decay_rate = 0.01
```

Step 4 Q-Learning algorithm

Write the required Q-table update/learn equation

Epsilon-greedy policy

In [5]:
```
for episode in range(total_episodes):
    # reset the environment
    state = env.reset()
    step = 0
    done = False
    total_training_rewards = 0
    print("****************************************")
    print("EPISODE ", episode)
    for step in range(max_steps):
        # If this is not the first step within the episode
        if step > 0:
            # Update Q(s,a) := Q(s,a) + lr * (R(s,a) + gamma * max(Q(s',a')) - Q(s,a))
            Q[state, action] = Q[state, action] + learning_rate * (reward + gamma * np.max(Q[state,:]) - Q[state, action])
        # Choose an action
        action = np.argmax(Q[state,:])
        # Take the action and observe the outcome state and reward
        new_state, reward, done, info = env.step(action)
        # Add to the cumulative reward
        total_training_rewards += reward
        # Our new state is state
        state = new_state
        # If we've reached the terminal state, then episode has finished
        if done == True:
            print("Total reward: %d", total_training_rewards)
            print("Training finished.\n")
            break
        # Update epsilon = max_epsilon + (min_epsilon - max_epsilon) * e^(-decay_rate*step)
        epsilon = min_epsilon + (max_epsilon - min_epsilon)*(np.exp(-decay_rate*episode))
        print("Epsilon: ", epsilon)
```

```
****************************************
EPISODE  0
Epsilon:  0.9900498337491681
Epsilon:  0.9801987737096835
Epsilon:  0.9704455335485082
Epsilon:  0.9607894391523232
Epsilon:  0.9512294245007139
Epsilon:  0.9417645335842487
Epsilon:  0.9323938199959688
Epsilon:  0.9231163463866358
Epsilon:  0.9139311852712282
Epsilon:  0.9048374180359596
Epsilon:  0.8958341358396976
Epsilon:  0.8869204367291954
Epsilon:  0.8780954297305631
Epsilon:  0.8693584326049159
Epsilon:  0.8607079766865892
Epsilon:  0.8521437111332003
Epsilon:  0.8436648851904954
Epsilon:  0.8352702608190503
Epsilon:  0.8269592976517852
Epsilon:  0.8187307530779818
Epsilon:  0.8105843906318474
Epsilon:  0.8025192762186115
Epsilon:  0.7945342761507115
Epsilon:  0.7866278610665535
Epsilon:  0.7787994062742436
Epsilon:  0.7710478206143685
Epsilon:  0.7633718788132373
Epsilon:  0.7557702408167954
Epsilon:  0.7482413966267542
Epsilon:  0.7407840167938307
Epsilon:  0.7333967391304348
Epsilon:  0.7260783436482085
Epsilon:  0.7188273391812865
Epsilon:  0.7116425986962552
Epsilon:  0.7045225602527812
Epsilon:  0.6974655055878928
Epsilon:  0.6904694224644992
Epsilon:  0.6835327316360353
Epsilon:  0.6766538164251207
Epsilon:  0.6698309871251556
Epsilon:  0.6630626098294973
Epsilon:  0.6563470031545941
Epsilon:  0.6496824871850499
Epsilon:  0.6430674356039386
Epsilon:  0.6365001307330195
Epsilon:  0.6299788417346094
Epsilon:  0.6235018750551276
Epsilon:  0.6170675986102862
Epsilon:  0.6106743233400561
Epsilon:  0.6043203933142566
Epsilon:  0.5980040097559426
Epsilon:  0.5917234206302211
Epsilon:  0.5854768154019385
Epsilon:  0.5792625743750423
Epsilon:  0.5730790762516655
Epsilon:  0.5669247233817701
Epsilon:  0.5607978409598503
Epsilon:  0.5546967670011148
Epsilon:  0.5486198426271889
Epsilon:  0.5425654206432022
Epsilon:  0.5365317776548671
Epsilon:  0.5305172274749322
Epsilon:  0.5245200876424189
Epsilon:  0.5185386421993079
Epsilon:  0.5125711548769414
Epsilon:  0.5066159867857991
Epsilon:  0.5006714954861408
Epsilon:  0.4947360206408523
Epsilon:  0.4888078977332079
Epsilon:  0.4828854314574899
Epsilon:  0.4769668942984276
Epsilon:  0.4710506431213077
Epsilon:  0.4651350948233383
Epsilon:  0.4592186489346525
Epsilon:  0.4533000398334688
Epsilon:  0.4473780788177339
Epsilon:  0.4414516375177344
Epsilon:  0.4355196342852819
Epsilon:  0.4295810591856784
Epsilon:  0.4236349178330095
Epsilon:  0.4176803339517026
Epsilon:  0.4117164339166702
Epsilon:  0.4057424386167896
Epsilon:  0.3997575351026692
Epsilon:  0.3937609152188296
Epsilon:  0.3877518359413202
Epsilon:  0.3817296098764732
Epsilon:  0.3756935395573835
Epsilon:  0.3696430252828761
Epsilon:  0.3635774420964037
Epsilon:  0.3574962898414056
Epsilon:  0.3513990094877594
Epsilon:  0.3452850190414423
Epsilon:  0.3391537911488641
Epsilon:  0.3330048276087632
Epsilon:  0.3268376446154074
Epsilon:  0.3206517234351274
Epsilon:  0.3144465823465118
Epsilon:  0.3082217185958254
Epsilon:  0.3019766184773780
Epsilon:  0.2957107821353831
Epsilon:  0.2894237666963690
Epsilon:  0.2831151981029494
Epsilon:  0.2767846193862718
Epsilon:  0.2704315985352862
Epsilon:  0.2640557344883757
Epsilon:  0.2576565411291846
Epsilon:  0.2512335329949238
Epsilon:  0.2447862777886835
Epsilon:  0.2383142871964028
Epsilon:  0.2318171777110094
Epsilon:  0.2252944496633959
Epsilon:  0.2187455968688845
Epsilon:  0.2121701417848953
Epsilon:  0.2055675279233116
Epsilon:  0.1989372206965967
Epsilon:  0.1922787192118226
Epsilon:  0.1855915217599993
Epsilon:  0.1788751668892508
Epsilon:  0.1721291542573386
Epsilon:  0.1653529296743840
Epsilon:  0.1585459438316371
Epsilon:  0.1517076562589814
Epsilon:  0.1448375268092980
Epsilon:  0.1379350206461830
Epsilon:  0.1309995635863374
Epsilon:  0.1240306403106808
Epsilon:  0.1170277824645101
Epsilon:  0.1099904988123515
Epsilon:  0.1029182436630035
Epsilon:  0.0958105268190750
Epsilon:  0.0886667155414353
Epsilon:  0.0814863103968477
Epsilon:  0.0742688117066091
Epsilon:  0.0670137680128428
Epsilon:  0.0597206704234199
Epsilon:  0.0523890267276166
Epsilon:  0.0450183274975335
Epsilon:  0.0376080390166959
Epsilon:  0.0301576438803403
Epsilon:  0.0226665659817503
Epsilon:  0.0151341077954261
Epsilon:  0.0075600172854922
Epsilon:  0.0000000000000000
Total reward: %d 0.0
Training finished.

****************************************
EPISODE  1
Epsilon:  0.9900498337491681
Epsilon:  0.9801987737096835
Epsilon:  0.9704455335485082
Epsilon:  0.9607894391523232
Epsilon:  0.9512294245007139
Epsilon:  0.9417645335842487
Epsilon:  0.9323938199959688
Epsilon:  0.9231163463866358
Epsilon:  0.9139311852712282
Epsilon:  0.9048374180359596
Epsilon:  0.8958341358396976
Epsilon:  0.8869204367291954
Epsilon:  0.8780954297305631
Epsilon:  0.8693584326049159
Epsilon:  0.8607079766865892
Epsilon:  0.8521437111332003
Epsilon:  0.8436648851904954
Epsilon:  0.8352702608190503
Epsilon:  0.8269592976517852
Epsilon:  0.8187307530779818
Epsilon:  0.8105843906318474
Epsilon:  0.8025192762186115
Epsilon:  0.7945342761507115
Epsilon:  0.7866278610665535
Epsilon:  0.7787994062742436
Epsilon:  0.7710478206143685
Epsilon:  0.7633718788132373
Epsilon:  0.7557702408167954
Epsilon:  0.7482413966267542
Epsilon:  0.7407840167938307
Epsilon:  0.7333967391304348
Epsilon:  0.7260783436482085
Epsilon:  0.7188273391812865
Epsilon:  0.7116425986962552
Epsilon:  0.7045225602527812
Epsilon:  0.6974655055878928
Epsilon:  0.6904694224644992
Epsilon:  0.6835327316360353
Epsilon:  0.6766538164251207
Epsilon:  0.6698309871251556
Epsilon:  0.6630626098294973
Epsilon:  0.6563470031545941
Epsilon:  0.6496824871850499
Epsilon:  0.6430674356039386
Epsilon:  0.6365001307330195
Epsilon:  0.6299788417346094
Epsilon:  0.6235018750551276
Epsilon:  0.6170675986102862
Epsilon:  0.6106743233400561
Epsilon:  0.6043203933142566
Epsilon:  0.5980040097559426
Epsilon:  0.5917234206302211
Epsilon:  0.5854768154019385
Epsilon:  0.5792625743750423
Epsilon:  0.5730790762516655
Epsilon:  0.5669247233817701
Epsilon:  0.5607978409598503
Epsilon:  0.5546967670011148
Epsilon:  0.5486198426271889
Epsilon:  0.5425654206432022
Epsilon:  0.5365317776548671
Epsilon:  0.5305172274749322
Epsilon:  0.5245200876424189
Epsilon:  0.5185386421993079
Epsilon:  0.5125711548769414
Epsilon:  0.5066159867857991
Epsilon:  0.5006714954861408
Epsilon:  0.4947360206408523
Epsilon:  0.4888078977332079
Epsilon:  0.4828854314574899
Epsilon:  0.4769668942984276
Epsilon:  0.4710506431213077
Epsilon:  0.4651350948233383
Epsilon:  0.4592186489346525
Epsilon:  0.4533000398334688
Epsilon:  0.4473780788177339
Epsilon:  0.4414516375177344
Epsilon:  0.4355196342852819
Epsilon:  0.4295810591856784
Epsilon:  0.4236349178330095
Epsilon:  0.4176803339517026
Epsilon:  0.4117164339166702
Epsilon:  0.4057424386167896
Epsilon:  0.3997575351026692
Epsilon:  0.3937609152188296
Epsilon:  0.3877518359413202
Epsilon:  0.3817296098764732
Epsilon:  0.3756935395573835
Epsilon:  0.3696430252828761
Epsilon:  0.3635774420964037
Epsilon:  0.3574962898414056
Epsilon:  0.3513990094877594
Epsilon:  0.3452850190414423
Epsilon:  0.3391537911488641
Epsilon:  0.3330048276087632
Epsilon:  0.3268376446154074
Epsilon:  0.3206517234351274
Epsilon:  0.3144465823465118
Epsilon:  0.3082217185958254
Epsilon:  0.3019766184773780
Epsilon:  0.2957107821353831
Epsilon:  0.2894237666963690
Epsilon:  0.2831151981029494
Epsilon:  0.2767846193862718
Epsilon:  0.2704315985352862
Epsilon:  0.2640557344883757
Epsilon:  0.2576565411291846
Epsilon:  0.2512335329949238
Epsilon:  0.2447862777886835
Epsilon:  0.2383142871964028
Epsilon:  0.2318171777110094
Epsilon:  0.2252944496633959
Epsilon:  0.2187455968688845
Epsilon:  0.2121701417848953
Epsilon:  0.2055675279233116
Epsilon:  0.1989372206965967
Epsilon:  0.1922787192118226
Epsilon:  0.1855915217599993
Epsilon:  0.1788751668892508
Epsilon:  0.1721291542573386
Epsilon:  0.1653529296743840
Epsilon:  0.1585459438316371
Epsilon:  0.1517076562589814
Epsilon:  0.1448375268092980
Epsilon:  0.1379350206461830
Epsilon:  0.1309995635863374
Epsilon:  0.1240306403106808
Epsilon:  0.1170277824645101
Epsilon:  0.1099904988123515
Epsilon:  0.1029182436630035
Epsilon:  0.0958105268190750
Epsilon:  0.0886667155414353
Epsilon:  0.0814863103968477
Epsilon:  0.0742688117066091
Epsilon:  0.0670137680128428
Epsilon:  0.0597206704234199
Epsilon:  0.0523890267276166
Epsilon:  0.0450183274975335
Epsilon:  0.0376080390166959
Epsilon:  0.0301576438803403
Epsilon:  0.0226665659817503
Epsilon:  0.0151341077954261
Epsilon:  0.0075600172854922
Epsilon:  0.0000000000000000
Total reward: %d 0.0
Training finished.

****************************************
EPISODE  2
Epsilon:  0.9900498337491681
Epsilon:  0.9801987737096835
Epsilon:  0.9704455335485082
Epsilon:  0.9607894391523232
Epsilon:  0.9512294245007139
Epsilon:  0.9417645335842487
Epsilon:  0.9323938199959688
Epsilon:  0.9231163463866358
Epsilon:  0.9139311852712282
Epsilon:  0.9048374180359596
Epsilon:  0.8958341358396976
Epsilon:  0.8869204367291954
Epsilon:  0.8780954297305631
Epsilon:  0.8693584326049159
Epsilon:  0.8607079766865892
Epsilon:  0.8521437111332003
Epsilon:  0.8436648851904954
Epsilon:  0.8352702608190503
Epsilon:  0.8269592976517852
Epsilon:  0.8187307530779818
Epsilon:  0.8105843906318474
Epsilon:  0.8025192762186115
Epsilon:  0.7945342761507115
Epsilon:  0.7866278610665535
Epsilon:  0.7787994062742436
Epsilon:  0.7710478206143685
Epsilon:  0.7633718788132373
Epsilon:  0.7557702408167954
Epsilon:  0.7482413966267542
Epsilon:  0.7407840167938307
Epsilon:  0.7333967391304348
Epsilon:  0.7260783436482085
Epsilon:  0.7188273391812865
Epsilon:  0.7116425986962552
Epsilon:  0.7045225602527812
Epsilon:  0.6974655055878928
Epsilon:  0.6904694224644992
Epsilon:  0.6835327316360353
Epsilon:  0.6766538164251207
Epsilon:  0.6698309871251556
Epsilon:  0.6630626098294973
Epsilon:  0.6563470031545941
Epsilon:  0.6496824871850499
Epsilon:  0.6430674356039386
Epsilon:  0.6365001307330195
Epsilon:  0.6299788417346094
Epsilon:  0.6235018750551276
Epsilon:  0.6170675986102862
Epsilon:  0.6106743233400561
Epsilon:  0.6043203933142566
Epsilon:  0.5980040097559426
Epsilon:  0.5917234206302211
Epsilon:  0.5854768154019385
Epsilon:  0.5792625743750423
Epsilon:  0.5730790762516655
Epsilon:  0.5669247233817701
Epsilon:  0.5607978409598503
Epsilon:  0.5546967670011148
Epsilon:  0.5486198426271889
Epsilon:  0.5425654206432022
Epsilon:  0.5365317776548671
Epsilon:  0.5305172274749322
Epsilon:  0.5245200876424189
Epsilon:  0.5185386421993079
Epsilon:  0.5125711548769414
Epsilon:  0.5066159867857991
Epsilon:  0.5006714954861408
Epsilon:  0.4947360206408523
Epsilon:  0.4888078977332079
Epsilon:  0.4828854314574899
Epsilon:  0.4769668942984276
Epsilon:  0.4710506431213077
Epsilon:  0.4651350948233383
Epsilon:  0.4592186489346525
Epsilon:  0.4533000398334688
Epsilon:  0.4473780788177339
Epsilon:  0.4414516375177344
Epsilon:  0.4355196342852819
Epsilon:  0.4295810591856784
Epsilon:  0.4236349178330095
Epsilon:  0.4176803339517026
Epsilon:  0.4117164339166702
Epsilon:  0.4057424386167896
Epsilon:  0.3997575351026692
Epsilon:  0.3937609152188296
Epsilon:  0.3877518359413202
Epsilon:  0.3817296098764732
Epsilon:  0.3756935395573835
Epsilon:  0.3696430252828761
Epsilon:  0.3635774420964037
Epsilon:  0.3574962898414056
Epsilon:  0.3513990094877594
Epsilon:  0.3452850190414423
Epsilon:  0.3391537911488641
Epsilon:  0.3330048276087632
Epsilon:  0.3268376446154074
Epsilon:  0.3206517234351274
Epsilon:  0.3144465823465118
Epsilon:  0.3082217185958254
Epsilon:  0.3019766184773780
Epsilon:  0.2957107821353831
Epsilon:  0.2894237666963690
Epsilon:  0.2831151981029494
Epsilon:  0.2767846193862718
Epsilon:  0.2704315985352862
Epsilon:  0.2640557344883757
Epsilon:  0.2576565411291846
Epsilon:  0.2512335329949238
Epsilon:  0.2447862777886835
Epsilon:  0.2383142871964028
Epsilon:  0.2318171777110094
Epsilon:  0.2252944496633959
Epsilon:  0.2187455968688845
Epsilon:  0.2121701417848953
Epsilon:  0.2055675279233116
Epsilon:  0.1989372206965967
Epsilon:  0.1922787192118226
Epsilon:  0.1855915217599993
Epsilon:  0.1788751668892508
Epsilon:  0.1721291542573386
Epsilon:  0.1653529296743840
Epsilon:  0.1585459438316371
Epsilon:  0.1517076562589814
Epsilon:  0.1448375268092980
Epsilon:  0.1379350206461830
Epsilon:  0.1309995635863374
Epsilon:  0.1240306403106808
Epsilon:  0.1170277824645101
Epsilon:  0.1099904988123515
Epsilon:  0.1029182436630035
Epsilon:  0.0958105268190750
Epsilon:  0.0886667155414353
Epsilon:  0.0814863103968477
Epsilon:  0.0742688117066091
Epsilon:  0.0670137680128428
Epsilon:  0.0597206704234199
Epsilon:  0.0523890267276166
Epsilon:  0.0450183274975335
Epsilon:  0.0376080390166959
Epsilon:  0.0301576438803403
Epsilon:  0.0226665659817503
Epsilon:  0.0151341077954261
Epsilon:  0.0075600172854922
Epsilon:  0.0000000000000000
Total reward: %d 0.0
Training finished.

****************************************
EPISODE  3
Epsilon:  0.9900498337491681
Epsilon:  0.9801987737096835
Epsilon:  0.9704455335485082
Epsilon:  0.9607894391523232
Epsilon:  0.9512294245007139
Epsilon:  0.9417645335842487
Epsilon:  0.9323938199959688
Epsilon:  0.9231163463866358
Epsilon:  0.9139311852712282
Epsilon:  0.9048374180359596
Epsilon:  0.8958341358396976
Epsilon:  0.8869204367291954
Epsilon:  0.8780954297305631
Epsilon:  0.8693584326049159
Epsilon:  0.8607079766865892
Epsilon:  0.8521437111332003
Epsilon:  0.8436648851904954
Epsilon:  0.8352702608190503
Epsilon:  0.8269592976517852
Epsilon:  0.8187307530779818
Epsilon:  0.8105843906318474
Epsilon:  0.8025192762186115
Epsilon:  0.7945342761507115
Epsilon:  0.7866278610665535
Epsilon:  0.7787994062742436
Epsilon:  0.7710478206143685
Epsilon:  0.7633718788132373
Epsilon:  0.7557702408167954
Epsilon:  0.7482413966267542
Epsilon:  0.7407840167938307
Epsilon:  0.7333967391304348
Epsilon:  0.7260783436482085
Epsilon:  0.7188273391812865
Epsilon:  0.7116425986962552
Epsilon:  0.7045225602527812
Epsilon:  0.6974655055878928
Epsilon:  0.6904694224644992
Epsilon:  0.6835327316360353
Epsilon:  0.6766538164251207
Epsilon:  0.6698309871251556
Epsilon:  0.6630626098294973
Epsilon:  0.6563470031545941
Epsilon:  0.6496824871850499
Epsilon:  0.6430674356039386
Epsilon:  0.6365001307330195
Epsilon:  0.6299788417346094
Epsilon:  0.6235018750551276
Epsilon:  0.6170675986102862
Epsilon:  0.6106743233400561
Epsilon:  0.6043203933142566
Epsilon:  0.5980040097559426
Epsilon:  0.5917234206302211
Epsilon:  0.5854768154019385
Epsilon:  0.5792625743750423
Epsilon:  0.5730790762516655
Epsilon:  0.5669247233817701
Epsilon:  0.5607978409598503
Epsilon:  0.5546967670011148
Epsilon:  0.5486198426271889
Epsilon:  0.5425654206432022
Epsilon:  0.5365317776548671
Epsilon:  0.5305172274749322
Epsilon:  0.5245200876424189
Epsilon:  0.5185386421993079
Epsilon:  0.5125711548769414
Epsilon:  0.5066159867857991
Epsilon:  0.5006714954861408
Epsilon:  0.4947360206408523
Epsilon:  0.4888078977332079
Epsilon:  0.4828854314574899
Epsilon:  0.4769668942984276
Epsilon:  0.4710506431213077
Epsilon:  0.4651350948233383
Epsilon:  0.4592186489346525
Epsilon:  0.4533000398334688
Epsilon:  0.4473780788177339
Epsilon:  0.4414516375177344
Epsilon:  0.4355196342852819
Epsilon:  0.4295810591856784
Epsilon:  0.4236349178330095
Epsilon:  0.4176803339517026
Epsilon:  0.4117164339166702
Epsilon:  0.4057424386167896
Epsilon:  0.3997575351026692
Epsilon:  0.3937609152188296
Epsilon:  0.3877518359413202
Epsilon:  0.3817296098764732
Epsilon:  0.3756935395573835
Epsilon:  0.3696430252828761
Epsilon:  0.3635774420964037
Epsilon:  0.3574962898414056
Epsilon:  0.3513990094877594
Epsilon:  0.3452850190414423
Epsilon:  0.3391537911488641
Epsilon:  0.3330048276087632
Epsilon:  0.3268376446154074
Epsilon:  0.3206517234351274
Epsilon:  0.3144465823465118
Epsilon:  0.3082217185958254
Epsilon:  0.3019766184773780
Epsilon:  0.2957107821353831
Epsilon:  0.2894237666963690
Epsilon:  0.2831151981029494
Epsilon:  0.2767846193862718
Epsilon:  0.2704315985352862
Epsilon:  0.2640557344883757
Epsilon:  0.2576565411291846
Epsilon:  0.2512335329949238
Epsilon:  0.2447862777886835
Epsilon:  0.2383142871964028
Epsilon:  0.2318171777110094
Epsilon:  0.2252944496633959
Epsilon:  0.2187455968688845
Epsilon:  0.2121701417848953
Epsilon:  0.2055675279233116
Epsilon:  0.1989372206965967
Epsilon:  0.1922787192118226
Epsilon:  0.1855915217599993
Epsilon:  0.1788751668892508
Epsilon:  0.1721291542573386
Epsilon:  0.1653529296743840
Epsilon:  0.1585459438316371
Epsilon:  0.1517076562589814
Epsilon:  0.1448375268092980
Epsilon:  0.1379350206461830
Epsilon:  0.1309995635863374
Epsilon:  0.1240306403106808
Epsilon:  0.1170277824645101
Epsilon:  0.1099904988123515
Epsilon:  0.1029182436630035
Epsilon:  0.0958105268190750
Epsilon:  0.0886667155414353
Epsilon:  0.0814863103968477
Epsilon:  0.0742688117066091
Epsilon:  0.0670137680128428
Epsilon:  0.0597206704234199
Epsilon:  0.0523890267276166
Epsilon:  0.0450183274975335
Epsilon:  0.0376080390166959
Epsilon:  0.0301576438803403
Epsilon:  0.0226665659817503
Epsilon:  0.0151341077954261
Epsilon:  0.0075600172854922
Epsilon:  0.0000000000000000
Total reward: %d 0.0
Training finished.

****************************************
EPISODE  4
Epsilon:  0.9900498337491681
Epsilon:  0.9801987737096835
Epsilon:  0.9704455335485082
Epsilon:  0.9607894391523232
Epsilon:  0.9512294245007139
Epsilon:  0.9417645335842487
Epsilon:  0.9323938199959688
Epsilon:  0.9231163463866358
Epsilon:  0.9139311852712282
Epsilon:  0.9048374180359596
Epsilon:  0.8958341358396976
Epsilon:  0.8869204367291954
Epsilon:  0.8780954297305631
Epsilon:  0.8693584326049159
Epsilon:  0.8607079766865892
Epsilon:  0.8521437111332003
Epsilon:  0.8436648851904954
Epsilon:  0.8352702608190503
Epsilon:  0.8269592976517852
Epsilon:  0.8187307530779818
Epsilon:  0.8105843906318474
Epsilon:  0.8025192762186115
Epsilon:  0.7945342761507115
Epsilon:  0.7866278610665535
Epsilon:  0.7787994062742436
Epsilon:  0.7710478206143685
Epsilon:  0.7633718788132373
Epsilon:  0.7557702408167954
Epsilon:  0.7482413966267542
Epsilon:  0.7407840167938307
Epsilon:  0.7333967391304348
Epsilon:  0.7260783436482085
Epsilon:  0.7188273391812865
Epsilon:  0.7116425986962552
Epsilon:  0.7045225602527812
Epsilon:  0.6974655055878928
Epsilon:  0.6904694224644992
Epsilon:  0.6835327316360353
Epsilon:  0.6766538164251207
Epsilon:  0.6698309871251556
Epsilon:  0.6630626098294973
Epsilon:  0.6563470031545941
Epsilon:  0.6496824871850499
Epsilon:  0.6430674356039386
Epsilon:  0.6365001307330195
Epsilon:  0.6299788417346094
Epsilon:  0.6235018750551276
Epsilon:  0.6170675986102862
Epsilon:  0.6106743233400561
Epsilon:  0.6043203933142566
Epsilon:  0.5980040097559426
Epsilon:  0.5917234206302211
Epsilon:  0.5854768154019385
Epsilon:  0.5792625743750423
Epsilon:  0.5730790762516655
Epsilon:  0.5669247233817701
Epsilon:  0.5607978409598503
Epsilon:  0.5546967670011148
Epsilon:  0.5486198426271889
Epsilon:  0.5425654206432022
Epsilon:  0.5365317776548671
Epsilon:  0.5305172274749322
Epsilon:  0.5245200876424189
Epsilon:  0.5185386421993079
Epsilon:  0.5125711548769414
Epsilon:  0.5066159867857991
Epsilon:  0.5006714954861408
Epsilon:  0.4947360206408523
Epsilon:  0.4888078977332079
Epsilon:  0.4828854314574899
Epsilon:  0.4769668942984276
Epsilon:  0.4710506431213077
Epsilon:  0.4651350948233383
Epsilon:  0.4592186489346525
Epsilon:  0.4533000398334688
Epsilon:  0.4473780788177339
Epsilon:  0.4414516375177344
Epsilon:  0.4355196342852819
Epsilon:  0.4295810591856784
Epsilon:  0.4236349178330095
Epsilon:  0.4176803339517026
Epsilon:  0.4117164339166702
Epsilon:  0.4057424386167896
Epsilon:  0.3997575351026692
Epsilon:  0.3937609152188296
Epsilon:  0.3877518359413202
Epsilon:  0.3817296098764732
Epsilon:  0.3756935395573835
Epsilon:  0.3696430252828761
Epsilon:  0.3635774420964037
Epsilon:  0.3574962898414056
Epsilon:  0.3513990094877594
Epsilon:  0.3452850190414423
Epsilon:  0.3391537911488641
Epsilon:  0.3330048276087632
Epsilon:  0.3268376446154074
Epsilon:  0.3206517234351274
Epsilon:  0.3144465823465118
Epsilon:  0.3082217185958254
Epsilon:  0.3019766184773780
Epsilon:  0.2957107821353831
Epsilon:  0.2894237666963690
Epsilon:  0.2831151981029494
Epsilon:  0.2767846193862718
Epsilon:  0.2704315985352862
Epsilon:  0.2640557344883757
Epsilon:  0.2576565411291846
Epsilon:  0.2512335329949238
Epsilon:  0.2447862777886835
Epsilon:  0.2383142871964028
Epsilon:  0.2318171777110094
Epsilon:  0.2252944496633959
Epsilon:  0.2187455968688845
Epsilon:  0.2121701417848953
Epsilon:  0.2055675279233116
Epsilon:  0.1989372206965967
Epsilon:  0.1922787192118226
Epsilon:  0.1855915217599993
Epsilon:  0.1788751668892508
Epsilon:  0.1721291542573386
Epsilon:  0.1653529296743840
Epsilon:  0.1585459438316371
Epsilon:  0.1517076562589814
Epsilon:  0.1448375268092980
Epsilon:  0.1379350206461830
Epsilon:  0.1309995635863374
Epsilon:  0.1240306403106808
Epsilon:  0.1170277824645101
Epsilon:  0.1099904988123515
Epsilon:  0.1029182436630035
Epsilon:  0.0958105268190750
Epsilon:  0.0886667155414353
Epsilon:  0.0814863103968477
Epsilon:  0.0742688117066091
Epsilon:  0.0670137680128428
Epsilon:  0.0597206704234199
Epsilon:  0.0523890267276166
Epsilon:  0.0450183274975335
Epsilon:  0.0376080390166959
Epsilon:  0.0301576438803403
Epsilon:  0.0226665659817503
Epsilon:  0.0151341077954261
Epsilon:  0.0075600172854922
Epsilon:  0.0000000000000000
Total reward: %d 0.0
Training finished.
```

Step 5 Test your Q-table

In [6]:
```
# Clear output
%clear_all
print("Average reward over 100 episodes:",np.sum(total_rewards)/5)
```

```
Average reward over 100 episodes: 0.0
```

| | | |

Show image as JSON decode/URL

© 2024 GitHub, Inc.