NAME : VAIBHAV BILOTIA

ROLL NO : 001811001036

DEPARTMENT : INFORMATION TECHNOLOGY

MACHINE LEARNING LAB

ASSIGNMENT - 5

GitHub Link: Link

Imports

In [ ]:
```
!pip install --no-cache gym[all]
!pip install IPython
!pip install Box2D
```

```
Requirement already satisfied: gym[all] in /usr/local/lib/python3.7/dist-packages
(0.17.3)
Requirement already satisfied: pyglet<=1.5.0,>=1.4.0 in /usr/local/lib/python3.7/dis
t-packages (from gym[all]) (1.5.0)
Requirement already satisfied: numpy>=1.10.4 in /usr/local/lib/python3.7/dist-packag
es (from gym[all]) (1.19.5)
Requirement already satisfied: cloudpickle<1.7.0,>=1.2.0 in /usr/local/lib/python3.
7/dist-packages (from gym[all]) (1.3.0)
Requirement already satisfied: scipy in /usr/local/lib/python3.7/dist-packages (from
gym[all]) (1.4.1)
Requirement already satisfied: opencv-python in /usr/local/lib/python3.7/dist-packag
es (from gym[all]) (4.1.2.30)
Requirement already satisfied: atari-py~=0.2.0 in /usr/local/lib/python3.7/dist-pack
ages (from gym[all]) (0.2.9)
Requirement already satisfied: imageio in /usr/local/lib/python3.7/dist-packages (fr
om gym[all]) (2.4.1)
Collecting box2d-py~=2.3.5
  Downloading box2d_py-2.3.8-cp37-cp37m-manylinux1_x86_64.whl (448 kB)
     |████████████████████████████████| 448 kB 4.1 MB/s
Collecting mujoco-py<2.0,>=1.50
  Downloading mujoco-py-1.50.1.68.tar.gz (120 kB)
     |████████████████████████████████| 120 kB 61.7 MB/s
Requirement already satisfied: Pillow in /usr/local/lib/python3.7/dist-packages (fro
m gym[all]) (7.1.2)
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from a
tari-py~=0.2.0->gym[all]) (1.15.0)
Collecting glfw>=1.4.0
  Downloading glfw-2.4.0-py2.py27.py3.py30.py31.py32.py33.py34.py35.py36.py37.py38-n
one-manylinux2014_x86_64.whl (205 kB)
     |████████████████████████████████| 205 kB 52.2 MB/s
Requirement already satisfied: Cython>=0.27.2 in /usr/local/lib/python3.7/dist-packa
ges (from mujoco-py<2.0,>=1.50->gym[all]) (0.29.24)
```

```
Requirement already satisfied: cffi>=1.10 in /usr/local/lib/python3.7/dist-packages
(from mujoco-py<2.0,>=1.50->gym[all]) (1.15.0)
Collecting lockfile>=0.12.2
  Downloading lockfile-0.12.2-py2.py3-none-any.whl (13 kB)
Requirement already satisfied: pycparser in /usr/local/lib/python3.7/dist-packages
(from cffi>=1.10->mujoco-py<2.0,>=1.50->gym[all]) (2.21)
Requirement already satisfied: future in /usr/local/lib/python3.7/dist-packages (fro
m pyglet<=1.5.0,>=1.4.0->gym[all]) (0.16.0)
Building wheels for collected packages: mujoco-py
  Building wheel for mujoco-py (setup.py) ... error
  ERROR: Failed building wheel for mujoco-py
  Running setup.py clean for mujoco-py
Failed to build mujoco-py
Installing collected packages: lockfile, glfw, mujoco-py, box2d-py
    Running setup.py install for mujoco-py ... error
ERROR: Command errored out with exit status 1: /usr/bin/python3 -u -c 'import io, o
s, sys, setuptools, tokenize; sys.argv[0] = '"'"'/tmp/pip-install-7uuogwls/mujoco-py
_d8fc0a1325eb4c7e9700987ee559cab2/setup.py'"'"'; __file__='"'"'/tmp/pip-install-7uuo
gwls/mujoco-py_d8fc0a1325eb4c7e9700987ee559cab2/setup.py'"'"';f = getattr(tokenize,
 '"'"'open'"'"', open)(__file__) if os.path.exists(__file__) else io.StringIO('"'"'f
rom setuptools import setup; setup()'"'"');code = f.read().replace('"'"'\r\n'"'"',
 '"'"'\n'"'"');f.close();exec(compile(code, __file__, '"'"'exec'"'"'))' install --re
cord /tmp/pip-record-n5xniv2r/install-record.txt --single-version-externally-managed
--compile --install-headers /usr/local/include/python3.7/mujoco-py Check the logs fo
r full command output.
Requirement already satisfied: IPython in /usr/local/lib/python3.7/dist-packages (5.
5.0)
Requirement already satisfied: pickleshare in /usr/local/lib/python3.7/dist-packages
(from IPython) (0.7.5)
Requirement already satisfied: decorator in /usr/local/lib/python3.7/dist-packages
 (from IPython) (4.4.2)
Requirement already satisfied: prompt-toolkit<2.0.0,>=1.0.4 in /usr/local/lib/python
3.7/dist-packages (from IPython) (1.0.18)
Requirement already satisfied: traitlets>=4.2 in /usr/local/lib/python3.7/dist-packa
ges (from IPython) (5.1.1)
Requirement already satisfied: simplegeneric>0.8 in /usr/local/lib/python3.7/dist-pa
ckages (from IPython) (0.8.1)
Requirement already satisfied: pygments in /usr/local/lib/python3.7/dist-packages (f
rom IPython) (2.6.1)
Requirement already satisfied: pexpect in /usr/local/lib/python3.7/dist-packages (fr
om IPython) (4.8.0)
Requirement already satisfied: setuptools>=18.5 in /usr/local/lib/python3.7/dist-pac
kages (from IPython) (57.4.0)
Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.7/dist-packages
 (from prompt-toolkit<2.0.0,>=1.0.4->IPython) (1.15.0)
Requirement already satisfied: wcwidth in /usr/local/lib/python3.7/dist-packages (fr
om prompt-toolkit<2.0.0,>=1.0.4->IPython) (0.2.5)
Requirement already satisfied: ptyprocess>=0.5 in /usr/local/lib/python3.7/dist-pack
ages (from pexpect->IPython) (0.7.0)
Collecting Box2D
  Downloading Box2D-2.3.10-cp37-cp37m-manylinux1_x86_64.whl (1.3 MB)
     |████████████████████████████████| 1.3 MB 3.7 MB/s
Installing collected packages: Box2D
Successfully installed Box2D-2.3.10
```

In [ ]:

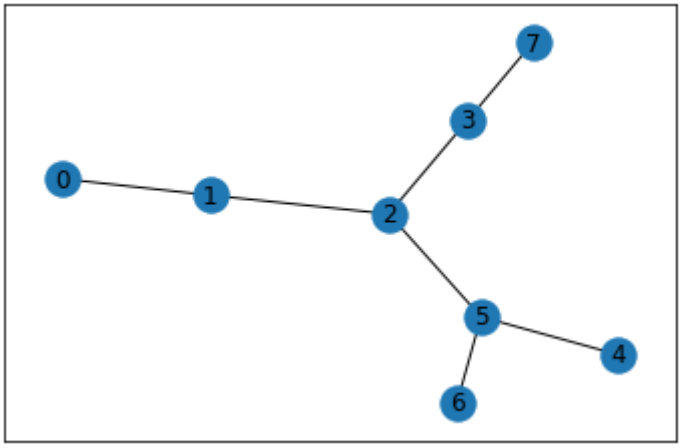
```
from __future__ import print_function
import os, sys, time, datetime, json, random
import numpy as np
from keras.models import Sequential
from keras.layers.core import Dense, Activation
from tensorflow.keras.optimizers import SGD , Adam, RMSprop
from keras.layers.advanced_activations import PReLU
import pylab as plt
import networkx as nx
from keras import models
from keras import layers
from collections import deque
import random
```

```
import gym
import pickle
from itertools import product

from matplotlib.pyplot import cm

from collections import defaultdict
```

Shortest Path Using Reinforcement Learning

In [ ]:
```
# map cell to cell, add circular cell to goal point
points_list = [(0,1), (2,5), (5,6), (4,5), (3,7), (2,3), (2,1)]
```

In [ ]:
```
goal = 7

G=nx.Graph()
G.add_edges_from(points_list)
pos = nx.spring_layout(G)
nx.draw_networkx_nodes(G,pos)
nx.draw_networkx_edges(G,pos)
nx.draw_networkx_labels(G,pos)
plt.show()
```



In [ ]:
```
# how many points in graph? x points
MATRIX_SIZE = 8

# create matrix x*y
R = np.matrix(np.ones(shape=(MATRIX_SIZE, MATRIX_SIZE)))
R *= -1
```

In [ ]:
```
# assign zeros to paths and 100 to goal-reaching point
for point in points_list:
    print(point)
    if point[1] == goal:
        R[point] = 100
    else:
        R[point] = 0

    if point[0] == goal:
        R[point[::-1]] = 100
    else:
```

```
        R[point[::-1]]= 0

R[goal,goal]= 100

print(R)
```

```
(0, 1)
(2, 5)
(5, 6)
(4, 5)
(3, 7)
(2, 3)
(2, 1)
[[ -1.   0.  -1.  -1.  -1.  -1.  -1.  -1.]
 [  0.  -1.   0.  -1.  -1.  -1.  -1.  -1.]
 [ -1.   0.  -1.   0.  -1.   0.  -1.  -1.]
 [ -1.  -1.   0.  -1.  -1.  -1.  -1. 100.]
 [ -1.  -1.  -1.  -1.  -1.   0.  -1.  -1.]
 [ -1.  -1.   0.  -1.   0.  -1.   0.  -1.]
 [ -1.  -1.  -1.  -1.  -1.   0.  -1.  -1.]
 [ -1.  -1.  -1.   0.  -1.  -1.  -1. 100.]]
```

In [ ]:

```
Q = np.matrix(np.zeros([MATRIX_SIZE,MATRIX_SIZE]))

# Learning parameter
gamma = 0.8

initial_state = 1

def available_actions(state):
    current_state_row = R[state,]
    av_act = np.where(current_state_row >= 0)[1]
    return av_act

available_act = available_actions(initial_state)

def sample_next_action(available_actions_range):
    next_action = int(np.random.choice(available_act,1))
    return next_action

action = sample_next_action(available_act)

def update(current_state, action, gamma):

  max_index = np.where(Q[action,] == np.max(Q[action,]))[1]

  if max_index.shape[0] > 1:
      max_index = int(np.random.choice(max_index, size = 1))
  else:
      max_index = int(max_index)
  max_value = Q[action, max_index]

  Q[current_state, action] = R[current_state, action] + gamma * max_value
  print('max_value', R[current_state, action] + gamma * max_value)

  if (np.max(Q) > 0):
    return(np.sum(Q/np.max(Q)*100))
  else:
    return (0)

update(initial_state, action, gamma)
```

```
max_value 0.0
```

Out[ ]:

```
0
```

In [ ]:

```
# Training
scores = []
for i in range(700):
    current_state = np.random.randint(0, int(Q.shape[0]))
    available_act = available_actions(current_state)
    action = sample_next_action(available_act)
    score = update(current_state,action,gamma)
    scores.append(score)
    print ('Score:', str(score))

print("Trained Q matrix:")
print(Q/np.max(Q)*100)
```

```
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
```

```
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 0.0
Score: 0
max_value 100.0
Score: 100.0
max_value 0.0
Score: 100.0
max_value 0.0
Score: 100.0
max_value 80.0
Score: 180.0
max_value 0.0
Score: 180.0
max_value 0.0
Score: 180.0
max_value 0.0
Score: 180.0
max_value 0.0
Score: 180.0
max_value 0.0
Score: 180.0
max_value 0.0
Score: 180.0
max_value 0.0
Score: 180.0
max_value 0.0
Score: 180.0
max_value 0.0
Score: 180.0
max_value 164.0
Score: 209.7560975609756
max_value 0.0
Score: 209.7560975609756
max_value 231.20000000000002
Score: 177.85467128027682
max_value 0.0
Score: 177.85467128027682
max_value 0.0
Score: 177.85467128027682
max_value 0.0
Score: 177.85467128027682
max_value 80.0
Score: 177.85467128027682
max_value 0.0
Score: 177.85467128027682
```

```
max_value 284.96000000000004
Score: 209.20830993823694
max_value 0.0
Score: 209.20830993823694
max_value 284.96000000000004
Score: 209.20830993823694
max_value 0.0
Score: 209.20830993823694
max_value 0.0
Score: 209.20830993823694
max_value 0.0
Score: 209.20830993823694
max_value 0.0
Score: 209.20830993823694
max_value 0.0
Score: 209.20830993823694
max_value 227.96800000000005
Score: 261.1341942728804
max_value 284.96000000000004
Score: 261.1341942728804
max_value 0.0
Score: 261.1341942728804
max_value 0.0
Score: 261.1341942728804
max_value 0.0
Score: 261.1341942728804
max_value 284.96000000000004
Score: 280.0
max_value 327.9680000000001
Score: 256.39574592643186
max_value 0.0
Score: 256.39574592643186
max_value 0.0
Score: 256.39574592643186
max_value 262.3744000000001
Score: 336.39574592643186
max_value 0.0
Score: 336.39574592643186
max_value 0.0
Score: 336.39574592643186
max_value 209.89952000000008
Score: 400.39574592643186
max_value 0.0
Score: 400.39574592643186
max_value 0.0
Score: 400.39574592643186
max_value 209.89952000000008
Score: 464.39574592643186
max_value 0.0
Score: 464.39574592643186
max_value 167.91961600000008
Score: 515.5957459264318
max_value 134.33569280000006
Score: 556.5557459264319
max_value 0.0
Score: 556.5557459264319
max_value 327.9680000000001
Score: 569.6692204117476
max_value 0.0
Score: 569.6692204117476
max_value 167.91961600000008
Score: 620.8692204117475
max_value 167.91961600000008
Score: 672.0692204117474
max_value 209.89952000000008
Score: 672.0692204117474
max_value 209.89952000000008
Score: 672.0692204117474
max_value 362.3744000000001
```

```
Score: 617.7529043994277
max_value 209.89952000000008
Score: 675.6762897158299
max_value 167.91961600000008
Score: 675.6762897158299
max_value 209.89952000000008
Score: 675.6762897158299
max_value 362.3744000000001
Score: 675.6762897158299
max_value 134.33569280000006
Score: 712.7472563183271
max_value 209.89952000000008
Score: 712.7472563183271
max_value 209.89952000000008
Score: 712.7472563183271
max_value 167.91961600000008
Score: 712.7472563183271
max_value 167.91961600000008
Score: 712.7472563183271
max_value 167.91961600000008
Score: 712.7472563183271
max_value 362.3744000000001
Score: 722.241966761449
max_value 134.33569280000006
Score: 759.3129333639463
max_value 209.89952000000008
Score: 759.3129333639463
max_value 167.91961600000008
Score: 759.3129333639463
max_value 209.89952000000008
Score: 759.3129333639463
max_value 134.33569280000006
Score: 759.3129333639463
max_value 167.91961600000008
Score: 805.651641617068
max_value 167.91961600000008
Score: 805.651641617068
max_value 209.89952000000008
Score: 805.651641617068
max_value 167.91961600000008
Score: 805.651641617068
max_value 167.91961600000008
Score: 805.651641617068
max_value 209.89952000000008
Score: 805.651641617068
max_value 167.91961600000008
Score: 851.9903498701897
max_value 167.91961600000008
Score: 851.9903498701897
max_value 167.91961600000008
Score: 851.9903498701897
max_value 209.89952000000008
Score: 851.9903498701897
max_value 167.91961600000008
Score: 851.9903498701897
max_value 209.89952000000008
Score: 851.9903498701897
max_value 167.91961600000008
Score: 851.9903498701897
max_value 209.89952000000008
Score: 851.9903498701897
max_value 167.91961600000008
Score: 851.9903498701897
max_value 167.91961600000008
Score: 851.9903498701897
max_value 289.8995200000001
Score: 859.5861182246871
max_value 167.91961600000008
Score: 859.5861182246871
```

```
max_value 389.8995200000001
Score: 805.9628179075471
max_value 167.91961600000008
Score: 805.9628179075471
max_value 167.91961600000008
Score: 805.9628179075471
max_value 167.91961600000008
Score: 805.9628179075471
max_value 389.8995200000001
Score: 805.9628179075471
max_value 134.33569280000006
Score: 805.9628179075471
max_value 167.91961600000008
Score: 805.9628179075471
max_value 167.91961600000008
Score: 805.9628179075471
max_value 231.9196160000001
Score: 811.6104514311791
max_value 167.91961600000008
Score: 811.6104514311791
max_value 389.8995200000001
Score: 811.6104514311791
max_value 167.91961600000008
Score: 811.6104514311791
max_value 167.91961600000008
Score: 811.6104514311791
max_value 167.91961600000008
Score: 811.6104514311791
max_value 134.33569280000006
Score: 811.6104514311791
max_value 134.33569280000006
Score: 811.6104514311791
max_value 311.91961600000013
Score: 817.258084954811
max_value 167.91961600000008
Score: 817.258084954811
max_value 311.91961600000013
Score: 817.258084954811
max_value 249.5356928000001
Score: 827.4238252973485
max_value 199.62855424000008
Score: 835.5564175713785
max_value 249.5356928000001
Score: 845.722157913916
max_value 199.62855424000008
Score: 845.722157913916
max_value 199.62855424000008
Score: 853.854750187946
max_value 159.7028433920001
Score: 860.36082400717
max_value 159.7028433920001
Score: 860.36082400717
max_value 249.5356928000001
Score: 860.36082400717
max_value 199.62855424000008
Score: 868.4934162812
max_value 389.8995200000001
Score: 875.5529581857398
max_value 199.62855424000008
Score: 875.5529581857398
max_value 199.62855424000008
Score: 875.5529581857398
max_value 249.5356928000001
Score: 875.5529581857398
max_value 199.62855424000008
Score: 875.5529581857398
max_value 199.62855424000008
Score: 875.5529581857398
max_value 199.62855424000008
```

```
Score: 875.5529581857398
max_value 199.62855424000008
Score: 875.5529581857398
max_value 411.91961600000013
Score: 834.0940183125439
max_value 199.62855424000008
Score: 841.79186444765
max_value 329.5356928000001
Score: 846.06844563382
max_value 411.91961600000013
Score: 846.06844563382
max_value 159.7028433920001
Score: 846.06844563382
max_value 199.62855424000008
Score: 846.06844563382
max_value 199.62855424000008
Score: 853.7662917689262
max_value 263.6285542400001
Score: 857.1875567178622
max_value 210.90284339200008
Score: 859.924568677011
max_value 159.7028433920001
Score: 866.0828455850959
max_value 263.6285542400001
Score: 866.0828455850959
max_value 159.7028433920001
Score: 872.2411224931808
max_value 263.6285542400001
Score: 872.2411224931808
max_value 199.62855424000008
Score: 872.2411224931808
max_value 159.7028433920001
Score: 872.2411224931808
max_value 159.7028433920001
Score: 872.2411224931808
max_value 199.62855424000008
Score: 872.2411224931808
max_value 199.62855424000008
Score: 872.2411224931808
max_value 199.62855424000008
Score: 872.2411224931808
max_value 199.62855424000008
Score: 872.2411224931808
max_value 159.7028433920001
Score: 872.2411224931808
max_value 199.62855424000008
Score: 872.2411224931808
max_value 263.6285542400001
Score: 872.2411224931808
max_value 168.7222747136001
Score: 874.4307320604998
max_value 263.6285542400001
Score: 882.128578195606
max_value 411.91961600000013
Score: 887.4743046783186
max_value 263.6285542400001
Score: 890.8955696272546
max_value 429.5356928000001
Score: 858.4594360791617
max_value 429.5356928000001
Score: 858.4594360791617
max_value 210.90284339200008
Score: 861.084197965306
max_value 263.6285542400001
Score: 861.084197965306
max_value 210.90284339200008
Score: 863.7089598514501
max_value 159.7028433920001
Score: 863.7089598514501
```

```
max_value 168.7222747136001
Score: 865.8087693603657
max_value 429.5356928000001
Score: 869.909959807466
max_value 159.7028433920001
Score: 869.909959807466
max_value 263.6285542400001
Score: 869.909959807466
max_value 168.7222747136001
Score: 869.909959807466
max_value 168.7222747136001
Score: 869.909959807466
max_value 343.6285542400001
Score: 896.8368383367093
max_value 159.7028433920001
Score: 896.8368383367093
max_value 210.90284339200008
Score: 899.4616002228536
max_value 263.6285542400001
Score: 899.4616002228536
max_value 210.90284339200008
Score: 899.4616002228536
max_value 168.7222747136001
Score: 899.4616002228536
max_value 210.90284339200008
Score: 899.4616002228536
max_value 168.7222747136001
Score: 899.4616002228536
max_value 210.90284339200008
Score: 899.4616002228536
max_value 210.90284339200008
Score: 902.0863621089977
max_value 168.7222747136001
Score: 904.1861716179133
max_value 168.7222747136001
Score: 904.1861716179133
max_value 168.7222747136001
Score: 904.1861716179133
max_value 210.90284339200008
Score: 904.1861716179133
max_value 210.90284339200008
Score: 904.1861716179133
max_value 210.90284339200008
Score: 904.1861716179133
max_value 210.90284339200008
Score: 904.1861716179133
max_value 210.90284339200008
Score: 904.1861716179133
max_value 263.6285542400001
Score: 904.1861716179133
max_value 263.6285542400001
Score: 904.1861716179133
max_value 263.6285542400001
Score: 904.1861716179133
max_value 210.90284339200008
Score: 904.1861716179133
max_value 210.90284339200008
Score: 904.1861716179133
max_value 443.6285542400001
Score: 878.6393843783252
max_value 263.6285542400001
Score: 878.6393843783252
max_value 210.90284339200008
Score: 878.6393843783252
max_value 210.90284339200008
Score: 878.6393843783252
max_value 210.90284339200008
Score: 878.6393843783252
max_value 263.6285542400001
```

```
Score: 878.6393843783252
max_value 210.90284339200008
Score: 878.6393843783252
max_value 168.7222747136001
Score: 878.6393843783252
max_value 210.90284339200008
Score: 878.6393843783252
max_value 168.7222747136001
Score: 878.6393843783252
max_value 343.6285542400001
Score: 878.6393843783252
max_value 263.6285542400001
Score: 878.6393843783252
max_value 210.90284339200008
Score: 878.6393843783252
max_value 343.6285542400001
Score: 881.8161098855783
max_value 168.7222747136001
Score: 881.8161098855783
max_value 343.6285542400001
Score: 881.8161098855783
max_value 210.90284339200008
Score: 881.8161098855783
max_value 210.90284339200008
Score: 881.8161098855783
max_value 210.90284339200008
Score: 881.8161098855783
max_value 274.9028433920001
Score: 884.3574902913807
max_value 454.9028433920001
Score: 868.0160074713311
max_value 210.90284339200008
Score: 868.0160074713311
max_value 210.90284339200008
Score: 868.0160074713311
max_value 210.90284339200008
Score: 868.0160074713311
max_value 210.90284339200008
Score: 868.0160074713311
max_value 210.90284339200008
Score: 868.0160074713311
max_value 219.92227471360007
Score: 869.9987234408218
max_value 210.90284339200008
Score: 869.9987234408218
max_value 219.92227471360007
Score: 871.9814394103123
max_value 210.90284339200008
Score: 871.9814394103123
max_value 210.90284339200008
Score: 871.9814394103123
max_value 210.90284339200008
Score: 871.9814394103123
max_value 168.7222747136001
Score: 871.9814394103123
max_value 219.92227471360007
Score: 873.964155379803
max_value 363.9222747136001
Score: 878.4252663111567
max_value 363.9222747136001
Score: 882.8863772425107
max_value 291.13781977088007
Score: 888.9336609494569
max_value 210.90284339200008
Score: 888.9336609494569
max_value 291.13781977088007
Score: 892.50254969454
max_value 363.9222747136001
Score: 892.50254969454
```

```
max_value 232.91025581670408
Score: 895.3576606906065
max_value 232.91025581670408
Score: 898.2127716866729
max_value 232.91025581670408
Score: 898.2127716866729
max_value 454.9028433920001
Score: 898.2127716866729
max_value 175.93781977088008
Score: 899.7989444622655
max_value 210.90284339200008
Score: 899.7989444622655
max_value 210.90284339200008
Score: 899.7989444622655
max_value 186.3282046533633
Score: 903.6692060347111
max_value 232.91025581670408
Score: 903.6692060347111
max_value 168.7222747136001
Score: 903.6692060347111
max_value 363.9222747136001
Score: 903.6692060347111
max_value 291.13781977088007
Score: 909.7164897416575
max_value 168.7222747136001
Score: 909.7164897416575
max_value 291.13781977088007
Score: 909.7164897416575
max_value 232.91025581670408
Score: 914.5543167072145
max_value 291.13781977088007
Score: 914.5543167072145
max_value 232.91025581670408
Score: 917.4094277032809
max_value 291.13781977088007
Score: 917.4094277032809
max_value 232.91025581670408
Score: 917.4094277032809
max_value 232.91025581670408
Score: 917.4094277032809
max_value 232.91025581670408
Score: 917.4094277032809
max_value 454.9028433920001
Score: 917.4094277032809
max_value 454.9028433920001
Score: 917.4094277032809
max_value 232.91025581670408
Score: 917.4094277032809
max_value 291.13781977088007
Score: 917.4094277032809
max_value 186.3282046533633
Score: 919.6935165001341
max_value 232.91025581670408
Score: 919.6935165001341
max_value 291.13781977088007
Score: 919.6935165001341
max_value 186.3282046533633
Score: 919.6935165001341
max_value 186.3282046533633
Score: 919.6935165001341
max_value 232.91025581670408
Score: 924.5313434656912
max_value 291.13781977088007
Score: 924.5313434656912
max_value 232.91025581670408
Score: 924.5313434656912
max_value 232.91025581670408
Score: 924.5313434656912
max_value 186.3282046533633
```

```
Score: 924.5313434656912
max_value 232.91025581670408
Score: 924.5313434656912
max_value 232.91025581670408
Score: 924.5313434656912
max_value 232.91025581670408
Score: 924.5313434656912
max_value 232.91025581670408
Score: 924.5313434656912
max_value 291.13781977088007
Score: 924.5313434656912
max_value 232.91025581670408
Score: 924.5313434656912
max_value 232.91025581670408
Score: 924.5313434656912
max_value 291.13781977088007
Score: 924.5313434656912
max_value 232.91025581670408
Score: 924.5313434656912
max_value 232.91025581670408
Score: 924.5313434656912
max_value 291.13781977088007
Score: 924.5313434656912
max_value 232.91025581670408
Score: 924.5313434656912
max_value 454.9028433920001
Score: 927.0097384275543
max_value 363.9222747136001
Score: 927.0097384275543
max_value 291.13781977088007
Score: 927.0097384275543
max_value 186.3282046533633
Score: 927.0097384275543
max_value 291.13781977088007
Score: 927.0097384275543
max_value 291.13781977088007
Score: 927.0097384275543
max_value 232.91025581670408
Score: 927.0097384275543
max_value 186.3282046533633
Score: 930.8800000000001
max_value 291.13781977088007
Score: 930.8800000000001
max_value 232.91025581670408
Score: 930.8800000000001
max_value 291.13781977088007
Score: 930.8800000000001
max_value 186.3282046533633
Score: 930.8800000000001
max_value 291.13781977088007
Score: 930.8800000000001
max_value 232.91025581670408
Score: 930.8800000000001
max_value 291.13781977088007
Score: 930.8800000000001
max_value 232.91025581670408
Score: 930.8800000000001
max_value 232.91025581670408
Score: 930.8800000000001
max_value 232.91025581670408
Score: 930.8800000000001
max_value 232.91025581670408
Score: 930.8800000000001
max_value 291.13781977088007
Score: 930.8800000000001
max_value 232.91025581670408
Score: 930.8800000000001
max_value 232.91025581670408
Score: 930.8800000000001
```

```
max_value 232.91025581670408
Score: 930.8800000000001
max_value 232.91025581670408
Score: 930.8800000000001
max_value 232.91025581670408
Score: 930.8800000000001
max_value 291.13781977088007
Score: 930.8800000000001
max_value 186.3282046533633
Score: 930.8800000000001
max_value 463.9222747136001
Score: 914.7262917066929
max_value 463.9222747136001
Score: 914.7262917066929
max_value 232.91025581670408
Score: 914.7262917066929
max_value 232.91025581670408
Score: 914.7262917066929
max_value 371.13781977088007
Score: 916.2816266088023
max_value 463.9222747136001
Score: 916.2816266088023
max_value 232.91025581670408
Score: 916.2816266088023
max_value 232.91025581670408
Score: 916.2816266088023
max_value 296.9102558167041
Score: 917.5258945304897
max_value 237.52820465336328
Score: 918.5213088678397
max_value 371.13781977088007
Score: 920.076643769949
max_value 371.13781977088007
Score: 920.076643769949
max_value 232.91025581670408
Score: 920.076643769949
max_value 237.52820465336328
Score: 921.0720581072989
max_value 296.9102558167041
Score: 922.3163260289863
max_value 237.52820465336328
Score: 922.3163260289863
max_value 296.9102558167041
Score: 923.5605939506737
max_value 463.9222747136001
Score: 925.5047625783104
max_value 296.9102558167041
Score: 925.5047625783104
max_value 237.52820465336328
Score: 926.5001769156604
max_value 237.52820465336328
Score: 926.5001769156604
max_value 237.52820465336328
Score: 926.5001769156604
max_value 237.52820465336328
Score: 926.5001769156604
max_value 237.52820465336328
Score: 926.5001769156604
max_value 237.52820465336328
Score: 926.5001769156604
max_value 190.02256372269062
Score: 927.2965083855403
max_value 237.52820465336328
Score: 927.2965083855403
max_value 237.52820465336328
Score: 927.2965083855403
max_value 237.52820465336328
Score: 928.2919227228901
max_value 296.9102558167041
```

```
Score: 928.2919227228901
max_value 237.52820465336328
Score: 928.2919227228901
max_value 237.52820465336328
Score: 928.2919227228901
max_value 237.52820465336328
Score: 928.2919227228901
max_value 190.02256372269062
Score: 928.2919227228901
max_value 237.52820465336328
Score: 928.2919227228901
max_value 237.52820465336328
Score: 928.2919227228901
max_value 190.02256372269062
Score: 928.2919227228901
max_value 296.9102558167041
Score: 928.2919227228901
max_value 471.13781977088007
Score: 915.6065100088471
max_value 237.52820465336328
Score: 915.6065100088471
max_value 190.02256372269062
Score: 916.3906455450526
max_value 296.9102558167041
Score: 916.3906455450526
max_value 371.13781977088007
Score: 916.3906455450526
max_value 237.52820465336328
Score: 916.3906455450526
max_value 296.9102558167041
Score: 916.3906455450526
max_value 476.9102558167041
Score: 906.509199859328
max_value 190.02256372269062
Score: 906.509199859328
max_value 371.13781977088007
Score: 906.509199859328
max_value 371.13781977088007
Score: 906.509199859328
max_value 237.52820465336328
Score: 906.509199859328
max_value 237.52820465336328
Score: 906.509199859328
max_value 237.52820465336328
Score: 906.509199859328
max_value 371.13781977088007
Score: 906.509199859328
max_value 237.52820465336328
Score: 906.509199859328
max_value 237.52820465336328
Score: 906.509199859328
max_value 296.9102558167041
Score: 906.509199859328
max_value 237.52820465336328
Score: 906.509199859328
max_value 237.52820465336328
Score: 906.509199859328
max_value 237.52820465336328
Score: 906.509199859328
max_value 237.52820465336328
Score: 906.509199859328
max_value 296.9102558167041
Score: 906.509199859328
max_value 237.52820465336328
Score: 906.509199859328
max_value 371.13781977088007
Score: 906.509199859328
max_value 237.52820465336328
Score: 906.509199859328
```

```
max_value 190.02256372269062
Score: 907.2838443598956
max_value 190.02256372269062
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 296.9102558167041
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 296.9102558167041
Score: 907.2838443598956
max_value 371.13781977088007
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 296.9102558167041
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 190.02256372269062
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 296.9102558167041
Score: 907.2838443598956
max_value 296.9102558167041
Score: 907.2838443598956
max_value 190.02256372269062
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 296.9102558167041
Score: 907.2838443598956
max_value 371.13781977088007
Score: 907.2838443598956
max_value 296.9102558167041
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 296.9102558167041
Score: 907.2838443598956
max_value 237.52820465336328
Score: 907.2838443598956
max_value 296.9102558167041
Score: 907.2838443598956
max_value 190.02256372269062
Score: 907.2838443598956
max_value 481.5282046533633
Score: 899.5418357840964
max_value 371.13781977088007
```

```
Score: 899.5418357840964
max_value 296.9102558167041
Score: 899.5418357840964
max_value 237.52820465336328
Score: 899.5418357840964
max_value 296.9102558167041
Score: 899.5418357840964
max_value 190.02256372269062
Score: 899.5418357840964
max_value 237.52820465336328
Score: 899.5418357840964
max_value 296.9102558167041
Score: 899.5418357840964
max_value 237.52820465336328
Score: 899.5418357840964
max_value 296.9102558167041
Score: 899.5418357840964
max_value 371.13781977088007
Score: 899.5418357840964
max_value 485.22256372269067
Score: 893.4543310941376
max_value 237.52820465336328
Score: 893.4543310941376
max_value 237.52820465336328
Score: 893.4543310941376
max_value 296.9102558167041
Score: 893.4543310941376
max_value 488.17805097815256
Score: 893.0138867480638
max_value 237.52820465336328
Score: 893.0138867480638
max_value 296.9102558167041
Score: 893.0138867480638
max_value 237.52820465336328
Score: 893.0138867480638
max_value 237.52820465336328
Score: 893.0138867480638
max_value 390.5424407825221
Score: 896.9887932340807
max_value 237.52820465336328
Score: 896.9887932340807
max_value 237.52820465336328
Score: 896.9887932340807
max_value 390.5424407825221
Score: 896.9887932340807
max_value 296.9102558167041
Score: 896.9887932340807
max_value 237.52820465336328
Score: 896.9887932340807
max_value 237.52820465336328
Score: 896.9887932340807
max_value 237.52820465336328
Score: 896.9887932340807
max_value 190.02256372269062
Score: 896.9887932340807
max_value 190.02256372269062
Score: 896.9887932340807
max_value 390.5424407825221
Score: 896.9887932340807
max_value 237.52820465336328
Score: 896.9887932340807
max_value 296.9102558167041
Score: 896.9887932340807
max_value 237.52820465336328
Score: 896.9887932340807
max_value 390.5424407825221
Score: 900.9636997200978
max_value 237.52820465336328
Score: 900.9636997200978
```

```
max_value 237.52820465336328
Score: 900.9636997200978
max_value 312.4339526260177
Score: 904.1436249089114
max_value 488.17805097815256
Score: 904.749036653569
max_value 390.5424407825221
Score: 904.749036653569
max_value 490.5424407825221
Score: 900.8701869167254
max_value 312.4339526260177
Score: 904.0347850266362
max_value 312.4339526260177
Score: 907.1993831365469
max_value 249.94716210081415
Score: 909.7310616244754
max_value 312.4339526260177
Score: 909.7310616244754
max_value 390.5424407825221
Score: 909.7310616244754
max_value 390.5424407825221
Score: 909.7310616244754
max_value 249.94716210081415
Score: 912.2627401124039
max_value 249.94716210081415
Score: 912.2627401124039
max_value 249.94716210081415
Score: 912.2627401124039
max_value 492.4339526260177
Score: 909.6228620016777
max_value 312.4339526260177
Score: 909.6228620016777
max_value 249.94716210081415
Score: 912.1448159370682
max_value 199.95772968065134
Score: 914.1623790853806
max_value 249.94716210081415
Score: 914.1623790853806
max_value 249.94716210081415
Score: 914.1623790853806
max_value 249.94716210081415
Score: 914.1623790853806
max_value 249.94716210081415
Score: 914.1623790853806
max_value 312.4339526260177
Score: 914.1623790853806
max_value 492.4339526260177
Score: 914.5464939160922
max_value 249.94716210081415
Score: 914.5464939160922
max_value 249.94716210081415
Score: 917.0684478514827
max_value 199.95772968065134
Score: 919.0860109997951
max_value 199.95772968065134
Score: 919.0860109997951
max_value 249.94716210081415
Score: 919.0860109997951
max_value 393.9471621008142
Score: 919.7774176950759
max_value 393.9471621008142
Score: 920.4688243903566
max_value 315.15772968065136
Score: 921.0219497465813
max_value 249.94716210081415
Score: 921.0219497465813
max_value 199.95772968065134
Score: 921.0219497465813
max_value 249.94716210081415
```

```
Score: 921.0219497465813
max_value 199.95772968065134
Score: 921.0219497465813
max_value 199.95772968065134
Score: 921.0219497465813
max_value 199.95772968065134
Score: 921.0219497465813
max_value 315.15772968065136
Score: 921.0219497465813
max_value 199.95772968065134
Score: 923.0395128948937
max_value 249.94716210081415
Score: 923.0395128948937
max_value 199.95772968065134
Score: 923.0395128948937
max_value 252.1261837445211
Score: 923.4820131798733
max_value 493.9471621008142
Score: 920.9592721252059
max_value 495.15772968065136
Score: 919.2577825165002
max_value 199.95772968065134
Score: 919.2577825165002
max_value 199.95772968065134
Score: 919.2577825165002
max_value 249.94716210081415
Score: 919.2577825165002
max_value 249.94716210081415
Score: 919.2577825165002
max_value 249.94716210081415
Score: 919.2577825165002
max_value 249.94716210081415
Score: 919.2577825165002
max_value 249.94716210081415
Score: 919.2577825165002
max_value 315.15772968065136
Score: 919.8078652257797
max_value 252.1261837445211
Score: 920.2479313932032
max_value 252.1261837445211
Score: 920.6879975606267
max_value 252.1261837445211
Score: 920.6879975606267
max_value 252.1261837445211
Score: 921.1280637280502
max_value 252.1261837445211
Score: 921.1280637280502
max_value 252.1261837445211
Score: 921.1280637280502
max_value 252.1261837445211
Score: 921.1280637280502
max_value 201.70094699561687
Score: 921.4801166619891
max_value 252.1261837445211
Score: 921.4801166619891
max_value 252.1261837445211
Score: 921.4801166619891
max_value 252.1261837445211
Score: 921.4801166619891
max_value 201.70094699561687
Score: 921.8321695959278
max_value 252.1261837445211
Score: 921.8321695959278
max_value 252.1261837445211
Score: 921.8321695959278
max_value 252.1261837445211
Score: 921.8321695959278
max_value 315.15772968065136
Score: 921.8321695959278
```

```
max_value 252.1261837445211
Score: 921.8321695959278
max_value 396.1261837445211
Score: 922.2722357633513
max_value 201.70094699561687
Score: 922.2722357633513
max_value 252.1261837445211
Score: 922.2722357633513
max_value 252.1261837445211
Score: 922.2722357633513
max_value 396.1261837445211
Score: 922.2722357633513
max_value 315.15772968065136
Score: 922.2722357633513
max_value 315.15772968065136
Score: 922.2722357633513
max_value 495.15772968065136
Score: 922.2722357633513
max_value 252.1261837445211
Score: 922.2722357633513
max_value 201.70094699561687
Score: 922.2722357633513
max_value 252.1261837445211
Score: 922.2722357633513
max_value 396.1261837445211
Score: 922.2722357633513
max_value 252.1261837445211
Score: 922.2722357633513
max_value 252.1261837445211
Score: 922.2722357633513
max_value 315.15772968065136
Score: 922.2722357633513
max_value 201.70094699561687
Score: 922.62428869729
max_value 315.15772968065136
Score: 922.62428869729
max_value 252.1261837445211
Score: 922.62428869729
max_value 315.15772968065136
Score: 922.62428869729
max_value 201.70094699561687
Score: 922.62428869729
max_value 315.15772968065136
Score: 923.1743714065694
max_value 252.1261837445211
Score: 923.1743714065694
max_value 315.15772968065136
Score: 923.1743714065694
max_value 315.15772968065136
Score: 923.1743714065694
max_value 252.1261837445211
Score: 923.1743714065694
max_value 252.1261837445211
Score: 923.1743714065694
max_value 495.15772968065136
Score: 923.1743714065694
max_value 252.1261837445211
Score: 923.1743714065694
max_value 252.1261837445211
Score: 923.1743714065694
max_value 252.1261837445211
Score: 923.1743714065694
max_value 252.1261837445211
Score: 923.1743714065694
max_value 201.70094699561687
Score: 923.1743714065694
max_value 201.70094699561687
Score: 923.1743714065694
max_value 252.1261837445211
```

```
Score: 923.1743714065694
max_value 252.1261837445211
Score: 923.1743714065694
max_value 252.1261837445211
Score: 923.1743714065694
max_value 252.1261837445211
Score: 923.1743714065694
max_value 315.15772968065136
Score: 923.1743714065694
max_value 495.15772968065136
Score: 923.4188526106936
max_value 252.1261837445211
Score: 923.4188526106936
max_value 315.15772968065136
Score: 923.4188526106936
max_value 315.15772968065136
Score: 923.4188526106936
max_value 496.1261837445211
Score: 921.8115128648672
max_value 201.70094699561687
Score: 921.8115128648672
max_value 396.1261837445211
Score: 921.8115128648672
max_value 201.70094699561687
Score: 921.8115128648672
max_value 396.1261837445211
Score: 921.8115128648672
max_value 496.9009469956169
Score: 920.5301521362454
max_value 315.15772968065136
Score: 920.5301521362454
max_value 497.5207575964935
Score: 919.507936924382
max_value 396.1261837445211
Score: 919.507936924382
max_value 315.15772968065136
Score: 919.507936924382
max_value 252.1261837445211
Score: 919.507936924382
max_value 252.1261837445211
Score: 919.507936924382
max_value 201.70094699561687
Score: 919.507936924382
max_value 252.1261837445211
Score: 919.507936924382
max_value 252.1261837445211
Score: 919.507936924382
max_value 252.1261837445211
Score: 919.507936924382
max_value 252.1261837445211
Score: 919.507936924382
max_value 252.1261837445211
Score: 919.507936924382
max_value 252.1261837445211
Score: 919.507936924382
max_value 201.70094699561687
Score: 919.507936924382
max_value 252.1261837445211
Score: 919.507936924382
max_value 315.15772968065136
Score: 919.507936924382
max_value 396.1261837445211
Score: 919.9459129505951
max_value 396.1261837445211
Score: 919.9459129505951
max_value 396.1261837445211
Score: 919.9459129505951
max_value 252.1261837445211
Score: 919.9459129505951
```

```
max_value 252.1261837445211
Score: 919.9459129505951
max_value 252.1261837445211
Score: 919.9459129505951
max_value 201.70094699561687
Score: 919.9459129505951
max_value 252.1261837445211
Score: 919.9459129505951
max_value 201.70094699561687
Score: 919.9459129505951
max_value 252.1261837445211
Score: 919.9459129505951
max_value 396.1261837445211
Score: 919.9459129505951
max_value 201.70094699561687
Score: 919.9459129505951
max_value 252.1261837445211
Score: 919.9459129505951
max_value 316.9009469956169
Score: 920.2962937715658
max_value 252.1261837445211
Score: 920.2962937715658
max_value 252.1261837445211
Score: 920.2962937715658
max_value 252.1261837445211
Score: 920.2962937715658
max_value 396.1261837445211
Score: 920.2962937715658
max_value 396.1261837445211
Score: 920.2962937715658
max_value 316.9009469956169
Score: 920.2962937715658
max_value 252.1261837445211
Score: 920.2962937715658
max_value 252.1261837445211
Score: 920.2962937715658
max_value 316.9009469956169
Score: 920.6466745925363
max_value 396.1261837445211
Score: 920.6466745925363
max_value 252.1261837445211
Score: 920.6466745925363
max_value 396.1261837445211
Score: 920.6466745925363
max_value 316.9009469956169
Score: 920.6466745925363
max_value 252.1261837445211
Score: 920.6466745925363
max_value 253.52075759649352
Score: 920.9269792493128
max_value 396.1261837445211
Score: 920.9269792493128
max_value 396.1261837445211
Score: 920.9269792493128
max_value 201.70094699561687
Score: 920.9269792493128
max_value 316.9009469956169
Score: 920.9269792493128
max_value 202.81660607719482
Score: 921.151222974734
max_value 316.9009469956169
Score: 921.151222974734
max_value 252.1261837445211
Score: 921.151222974734
max_value 253.52075759649352
Score: 921.4315276315103
max_value 253.52075759649352
Score: 921.4315276315103
max_value 252.1261837445211
```

```
Score: 921.4315276315103
max_value 252.1261837445211
Score: 921.4315276315103
max_value 252.1261837445211
Score: 921.4315276315103
max_value 252.1261837445211
Score: 921.4315276315103
max_value 252.1261837445211
Score: 921.4315276315103
max_value 316.9009469956169
Score: 921.4315276315103
max_value 202.81660607719482
Score: 921.4315276315103
max_value 252.1261837445211
Score: 921.4315276315103
max_value 316.9009469956169
Score: 921.4315276315103
max_value 316.9009469956169
Score: 921.4315276315103
max_value 316.9009469956169
Score: 921.4315276315103
max_value 498.01660607719487
Score: 920.6136722226626
max_value 252.1261837445211
Score: 920.6136722226626
max_value 498.4132848617559
Score: 920.5341550917176
max_value 201.70094699561687
Score: 920.5341550917176
max_value 253.52075759649352
Score: 920.5341550917176
max_value 252.1261837445211
Score: 920.5341550917176
max_value 252.1261837445211
Score: 920.5341550917176
max_value 316.9009469956169
Score: 920.5341550917176
max_value 253.52075759649352
Score: 920.5341550917176
max_value 202.81660607719482
Score: 920.5341550917176
max_value 253.52075759649352
Score: 920.5341550917176
max_value 252.1261837445211
Score: 920.5341550917176
max_value 252.1261837445211
Score: 920.5341550917176
max_value 201.70094699561687
Score: 920.5341550917176
max_value 253.52075759649352
Score: 920.5341550917176
max_value 253.52075759649352
Score: 920.5341550917176
max_value 202.81660607719482
Score: 920.5341550917176
max_value 316.9009469956169
Score: 920.5341550917176
max_value 252.1261837445211
Score: 920.5341550917176
max_value 202.81660607719482
Score: 920.5341550917176
max_value 252.1261837445211
Score: 920.5341550917176
max_value 498.4132848617559
Score: 920.6137434166296
max_value 316.9009469956169
Score: 920.6137434166296
max_value 316.9009469956169
Score: 920.6137434166296
```

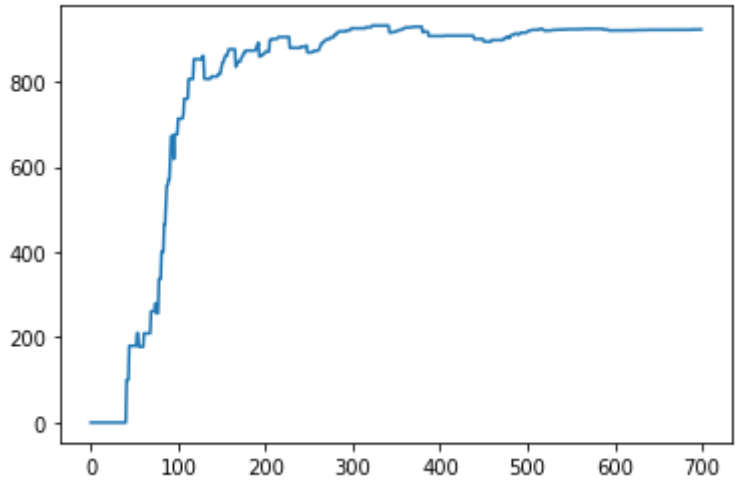
```
max_value 252.1261837445211
Score: 920.6137434166296
max_value 252.1261837445211
Score: 920.6137434166296
max_value 253.52075759649352
Score: 920.6137434166296
max_value 201.70094699561687
Score: 920.6137434166296
max_value 252.1261837445211
Score: 920.6137434166296
max_value 398.7306278894048
Score: 921.1362905123809
max_value 318.98450231152384
Score: 921.554328188982
max_value 318.98450231152384
Score: 921.972365865583
max_value 498.7306278894048
Score: 921.4493436474943
max_value 255.18760184921908
Score: 921.7835609903873
max_value 252.1261837445211
Score: 921.7835609903873
max_value 498.98450231152384
Score: 921.365452159979
max_value 255.18760184921908
Score: 921.365452159979
max_value 255.18760184921908
Score: 921.6994994590441
max_value 204.15008147937527
Score: 921.9667372982963
max_value 252.1261837445211
Score: 921.9667372982963
max_value 255.18760184921908
Score: 921.9667372982963
max_value 201.70094699561687
Score: 921.9667372982963
Trained Q matrix:
[[  0.          51.14138829   0.           0.           0.
    0.           0.           0.        ]
 [ 40.91311063   0.          63.92673537   0.           0.
    0.           0.           0.        ]
 [  0.          51.14138829   0.          79.90841921   0.
   46.67685163   0.           0.        ]
 [  0.           0.          63.92673537   0.           0.
    0.           0.          99.88552401]
 [  0.           0.           0.           0.           0.
   50.5278586    0.           0.        ]
 [  0.           0.          63.15982325   0.          40.42228688
    0.          40.42228688   0.        ]
 [  0.           0.           0.           0.           0.
   50.5278586    0.           0.        ]
 [  0.           0.           0.          79.3864703    0.
    0.           0.         100.        ]]
```

In [ ]:

```
current_state = 0
steps = [current_state]

while current_state != 7:

    next_step_index = np.where(Q[current_state,]
        == np.max(Q[current_state,]))[1]

    if next_step_index.shape[0] > 1:
        next_step_index = int(np.random.choice(next_step_index, size = 1))
    else:
        next_step_index = int(next_step_index)
```

```
    steps.append(next_step_index)
    current_state = next_step_index

print(f"Most efficient path: {steps}")
plt.plot(scores)
plt.show()
```

```
Most efficient path: [0, 1, 2, 3, 7]
```



Shortest Path Using Deep Reinforcement Learning

In [ ]:

```
import matplotlib.pyplot as plt
%matplotlib inline
```

In [ ]:

```
maze = np.array([
    [ 1.,  0.,  1.,  1.,  1.,  1.,  1.,  1.,  1.,  1.],
    [ 1.,  1.,  1.,  1.,  1.,  0.,  1.,  1.,  1.,  1.],
    [ 1.,  1.,  1.,  1.,  1.,  0.,  1.,  1.,  1.,  1.],
    [ 0.,  0.,  1.,  0.,  0.,  1.,  0.,  1.,  1.,  1.],
    [ 1.,  1.,  0.,  1.,  0.,  1.,  0.,  0.,  0.,  1.],
    [ 1.,  1.,  0.,  1.,  0.,  1.,  1.,  1.,  1.,  1.],
    [ 1.,  1.,  1.,  1.,  1.,  1.,  1.,  1.,  1.,  1.],
    [ 1.,  1.,  1.,  1.,  1.,  1.,  0.,  0.,  0.,  0.],
    [ 1.,  0.,  0.,  0.,  0.,  0.,  1.,  1.,  1.,  1.],
    [ 1.,  1.,  1.,  1.,  1.,  1.,  1.,  0.,  1.,  1.]
])
```

In [ ]:

```
visited_mark = 0.8  # Cells visited by the rat will be painted by grayscale value 0.
rat_mark = 0.5      # The current rat cell will be painteg by grayscale value 0.5
LEFT, UP, RIGHT, DOWN = 0, 1, 2, 3

# Actions memo
actions_dict = {
    LEFT: 'left',
    UP: 'up',
    RIGHT: 'right',
    DOWN: 'down',
}

num_actions = len(actions_dict)

epsilon = 0.1   # Exploration factor
```

In [ ]:

```
# maze is a 2d Numpy array of floats between 0.0 to 1.0
# 1.0 corresponds to a free cell, and 0.0 an occupied cell
# rat = (row, col) initial rat position (defaults to (0,0))

class Qmaze(object):
    def __init__(self, maze, rat=(0,0)):
        self._maze = np.array(maze)
        nrows, ncols = self._maze.shape
        self.target = (nrows-1, ncols-1)   # target cell where the "cheese" is
        self.free_cells = [(r,c) for r in range(nrows) for c in range(ncols) if self
        self.free_cells.remove(self.target)
        if self._maze[self.target] == 0.0:
            raise Exception("Invalid maze: target cell cannot be blocked!")
        if not rat in self.free_cells:
            raise Exception("Invalid Rat Location: must sit on a free cell")
        self.reset(rat)

    def reset(self, rat):
        self.rat = rat
        self.maze = np.copy(self._maze)
        nrows, ncols = self.maze.shape
        row, col = rat
        self.maze[row, col] = rat_mark
        self.state = (row, col, 'start')
        self.min_reward = -0.5 * self.maze.size
        self.total_reward = 0
        self.visited = set()

    def update_state(self, action):
        nrows, ncols = self.maze.shape
        nrow, ncol, nmode = rat_row, rat_col, mode = self.state

        if self.maze[rat_row, rat_col] > 0.0:
            self.visited.add((rat_row, rat_col))  # mark visited cell

        valid_actions = self.valid_actions()

        if not valid_actions:
            nmode = 'blocked'
        elif action in valid_actions:
            nmode = 'valid'
            if action == LEFT:
                ncol -= 1
            elif action == UP:
                nrow -= 1
            if action == RIGHT:
                ncol += 1
            elif action == DOWN:
                nrow += 1
        else:                  # invalid action, no change in rat position
            mode = 'invalid'

        # new state
        self.state = (nrow, ncol, nmode)

    def get_reward(self):
        rat_row, rat_col, mode = self.state
        nrows, ncols = self.maze.shape
        if rat_row == nrows-1 and rat_col == ncols-1:
            return 1.0
        if mode == 'blocked':
            return self.min_reward - 1
```

```
        if (rat_row, rat_col) in self.visited:
            return -0.25
        if mode == 'invalid':
            return -0.75
        if mode == 'valid':
            return -0.04

    def act(self, action):
        self.update_state(action)
        reward = self.get_reward()
        self.total_reward += reward
        status = self.game_status()
        envstate = self.observe()
        return envstate, reward, status

    def observe(self):
        canvas = self.draw_env()
        envstate = canvas.reshape((1, -1))
        return envstate

    def draw_env(self):
        canvas = np.copy(self.maze)
        nrows, ncols = self.maze.shape
        # clear all visual marks
        for r in range(nrows):
            for c in range(ncols):
                if canvas[r,c] > 0.0:
                    canvas[r,c] = 1.0
        # draw the rat
        row, col, valid = self.state
        canvas[row, col] = rat_mark
        return canvas

    def game_status(self):
        if self.total_reward < self.min_reward:
            return 'lose'
        rat_row, rat_col, mode = self.state
        nrows, ncols = self.maze.shape
        if rat_row == nrows-1 and rat_col == ncols-1:
            return 'win'

        return 'not_over'

    def valid_actions(self, cell=None):
        if cell is None:
            row, col, mode = self.state
        else:
            row, col = cell
        actions = [0, 1, 2, 3]
        nrows, ncols = self.maze.shape
        if row == 0:
            actions.remove(1)
        elif row == nrows-1:
            actions.remove(3)

        if col == 0:
            actions.remove(0)
        elif col == ncols-1:
            actions.remove(2)

        if row>0 and self.maze[row-1,col] == 0.0:
            actions.remove(1)
        if row<nrows-1 and self.maze[row+1,col] == 0.0:
            actions.remove(3)
```

```
        if col>0 and self.maze[row,col-1] == 0.0:
            actions.remove(0)
        if col<ncols-1 and self.maze[row,col+1] == 0.0:
            actions.remove(2)

        return actions
```

In [ ]:

```
def show(qmaze):
    plt.grid('on')
    nrows, ncols = qmaze.maze.shape
    ax = plt.gca()
    ax.set_xticks(np.arange(0.5, nrows, 1))
    ax.set_yticks(np.arange(0.5, ncols, 1))
    ax.set_xticklabels([])
    ax.set_yticklabels([])
    canvas = np.copy(qmaze.maze)
    for row,col in qmaze.visited:
        canvas[row,col] = 0.6
    rat_row, rat_col, _ = qmaze.state
    canvas[rat_row, rat_col] = 0.3   # rat cell
    canvas[nrows-1, ncols-1] = 0.9 # cheese cell
    img = plt.imshow(canvas, interpolation='none', cmap='gray')
    return img
```

In [ ]:

```
maze = [
    [ 1.,  0.,  1.,  1.,  1.,  1.,  1.,  1.],
    [ 1.,  0.,  1.,  1.,  1.,  0.,  1.,  1.],
    [ 1.,  1.,  1.,  1.,  0.,  1.,  0.,  1.],
    [ 1.,  1.,  1.,  0.,  1.,  1.,  1.,  1.],
    [ 1.,  1.,  0.,  1.,  1.,  1.,  1.,  1.],
    [ 1.,  1.,  1.,  0.,  1.,  0.,  0.,  0.],
    [ 1.,  1.,  1.,  0.,  1.,  1.,  1.,  1.],
    [ 1.,  1.,  1.,  1.,  0.,  1.,  1.,  1.]
]
```

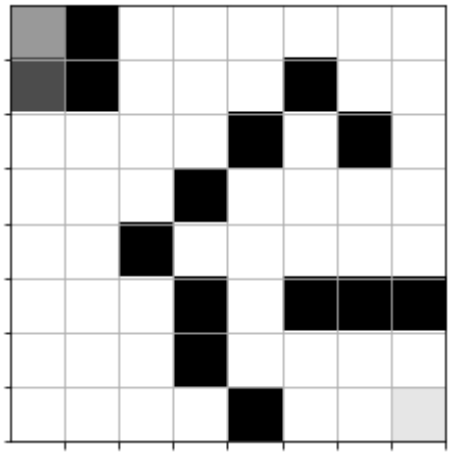
In [ ]:

```
qmaze = Qmaze(maze)
canvas, reward, game_over = qmaze.act(DOWN)
print("reward=", reward)
show(qmaze)
```

```
reward= -0.04
```

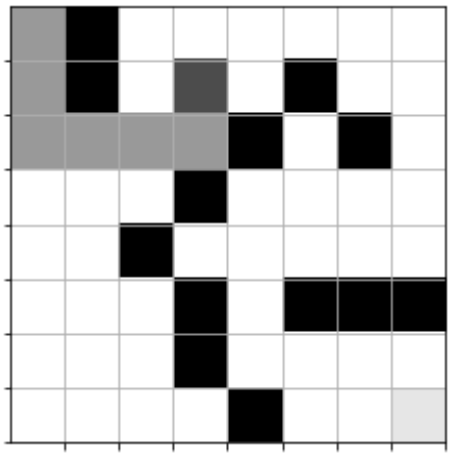
Out[ ]:

```
<matplotlib.image.AxesImage at 0x7fd01d782890>
```



In [ ]:
```
qmaze.act(DOWN)  # move down
qmaze.act(RIGHT)  # move right
qmaze.act(RIGHT)  # move right
qmaze.act(RIGHT)  # move right
qmaze.act(UP)  # move up
show(qmaze)
```

Out[ ]: <matplotlib.image.AxesImage at 0x7fd01d259c50>



In [ ]:
```
def play_game(model, qmaze, rat_cell):
    qmaze.reset(rat_cell)
    envstate = qmaze.observe()
    while True:
        prev_envstate = envstate
        # get next action
        q = model.predict(prev_envstate)
        action = np.argmax(q[0])

        # apply action, get rewards and new state
        envstate, reward, game_status = qmaze.act(action)
        if game_status == 'win':
            return True
        elif game_status == 'lose':
            return False
```

In [ ]:
```
def completion_check(model, qmaze):
    for cell in qmaze.free_cells:
        if not qmaze.valid_actions(cell):
            return False
        if not play_game(model, qmaze, cell):
            return False
    return True
```

In [ ]:
```
class Experience(object):
    def __init__(self, model, max_memory=100, discount=0.95):
        self.model = model
        self.max_memory = max_memory
        self.discount = discount
        self.memory = list()
        self.num_actions = model.output_shape[-1]

    def remember(self, episode):
        self.memory.append(episode)
        if len(self.memory) > self.max_memory:
            del self.memory[0]
```

```
    def predict(self, envstate):
        return self.model.predict(envstate)[0]

    def get_data(self, data_size=10):
        env_size = self.memory[0][0].shape[1]   # envstate 1d size (1st element of e
        mem_size = len(self.memory)
        data_size = min(mem_size, data_size)
        inputs = np.zeros((data_size, env_size))
        targets = np.zeros((data_size, self.num_actions))
        for i, j in enumerate(np.random.choice(range(mem_size), data_size, replace=F
            envstate, action, reward, envstate_next, game_over = self.memory[j]
            inputs[i] = envstate
            # There should be no target values for actions not taken.
            targets[i] = self.predict(envstate)
            # Q_sa = derived policy = max quality env/action = max_a' Q(s', a')
            Q_sa = np.max(self.predict(envstate_next))
            if game_over:
                targets[i, action] = reward
            else:
                # reward + gamma * max_a' Q(s', a')
                targets[i, action] = reward + self.discount * Q_sa
        return inputs, targets
```

In [ ]:

```
def qtrain(model, maze, **opt):
    global epsilon
    n_epoch = opt.get('n_epoch', 15000)
    max_memory = opt.get('max_memory', 1000)
    data_size = opt.get('data_size', 50)
    weights_file = opt.get('weights_file', "")
    name = opt.get('name', 'model')
    start_time = datetime.datetime.now()

    if weights_file:
        print("loading weights from file: %s" % (weights_file,))
        model.load_weights(weights_file)

    qmaze = Qmaze(maze)

    # Initialize experience replay object
    experience = Experience(model, max_memory=max_memory)

    win_history = []   # history of win/lose game
    n_free_cells = len(qmaze.free_cells)
    hsize = qmaze.maze.size//2   # history window size
    win_rate = 0.0
    imctr = 1

    for epoch in range(n_epoch):
        loss = 0.0
        rat_cell = random.choice(qmaze.free_cells)
        qmaze.reset(rat_cell)
        game_over = False

        # get initial envstate (1d flattened canvas)
        envstate = qmaze.observe()

        n_episodes = 0
        while not game_over:
            valid_actions = qmaze.valid_actions()
            if not valid_actions: break
            prev_envstate = envstate
            # Get next action
```

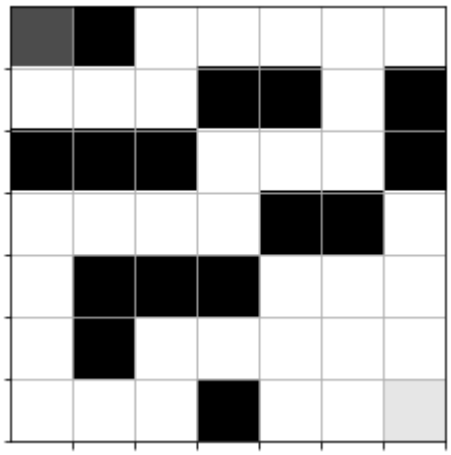
```
            if np.random.rand() < epsilon:
                action = random.choice(valid_actions)
            else:
                action = np.argmax(experience.predict(prev_envstate))

            # Apply action, get reward and new envstate
            envstate, reward, game_status = qmaze.act(action)
            if game_status == 'win':
                win_history.append(1)
                game_over = True
            elif game_status == 'lose':
                win_history.append(0)
                game_over = True
            else:
                game_over = False

            # Store episode (experience)
            episode = [prev_envstate, action, reward, envstate, game_over]
            experience.remember(episode)
            n_episodes += 1

            # Train neural network model
            inputs, targets = experience.get_data(data_size=data_size)
            h = model.fit(
                inputs,
                targets,
                epochs=8,
                batch_size=16,
                verbose=0,
            )
            loss = model.evaluate(inputs, targets, verbose=0)

        if len(win_history) > hsize:
            win_rate = sum(win_history[-hsize:]) / hsize

        dt = datetime.datetime.now() - start_time
        t = format_time(dt.total_seconds())
        template = "Epoch: {:03d}/{:d} | Loss: {:.4f} | Episodes: {:d} | Win count:
        print(template.format(epoch, n_epoch-1, loss, n_episodes, sum(win_history),
        # we simply check if training has exhausted all free cells and if in all
        # cases the agent won
        if win_rate > 0.9 : epsilon = 0.05
        if sum(win_history[-hsize:]) == hsize and completion_check(model, qmaze):
            print("Reached 100%% win rate at epoch: %d" % (epoch,))
            break

    # Save trained model weights and architecture, this will be used by the visualiz
    h5file = name + ".h5"
    json_file = name + ".json"
    model.save_weights(h5file, overwrite=True)
    with open(json_file, "w") as outfile:
        json.dump(model.to_json(), outfile)
    end_time = datetime.datetime.now()
    dt = datetime.datetime.now() - start_time
    seconds = dt.total_seconds()
    t = format_time(seconds)
    print('files: %s, %s' % (h5file, json_file))
    print("n_epoch: %d, max_mem: %d, data: %d, time: %s" % (epoch, max_memory, data_
    return seconds

# This is a small utility for printing readable time strings:
def format_time(seconds):
    if seconds < 400:
        s = float(seconds)
```

```
        return "%.1f seconds" % (s,)
    elif seconds < 4000:
        m = seconds / 60.0
        return "%.2f minutes" % (m,)
    else:
        h = seconds / 3600.0
        return "%.2f hours" % (h,)
```

In [ ]:
```
def build_model(maze, lr=0.001):
    model = Sequential()
    model.add(Dense(maze.size, input_shape=(maze.size,)))
    model.add(PReLU())
    model.add(Dense(maze.size))
    model.add(PReLU())
    model.add(Dense(num_actions))
    model.compile(optimizer='adam', loss='mse')
    return model
```

In [ ]:
```
maze =  np.array([
    [ 1.,  0.,  1.,  1.,  1.,  1.,  1.],
    [ 1.,  1.,  1.,  0.,  0.,  1.,  0.],
    [ 0.,  0.,  0.,  1.,  1.,  1.,  0.],
    [ 1.,  1.,  1.,  1.,  0.,  0.,  1.],
    [ 1.,  0.,  0.,  0.,  1.,  1.,  1.],
    [ 1.,  0.,  1.,  1.,  1.,  1.,  1.],
    [ 1.,  1.,  1.,  0.,  1.,  1.,  1.]
])

qmaze = Qmaze(maze)
show(qmaze)
```

Out[ ]: <matplotlib.image.AxesImage at 0x7fd01d1dda10>



In [ ]:
```
model = build_model(maze)
qtrain(model, maze, epochs=100, max_memory=8*maze.size, data_size=32)
```

```
Epoch: 000/14999 | Loss: 0.0023 | Episodes: 106 | Win count: 0 | Win rate: 0.000 | t
ime: 302.9 seconds
Epoch: 001/14999 | Loss: 0.0565 | Episodes: 111 | Win count: 0 | Win rate: 0.000 | t
ime: 11.24 minutes
Epoch: 002/14999 | Loss: 0.0019 | Episodes: 104 | Win count: 0 | Win rate: 0.000 | t
ime: 17.00 minutes
Epoch: 003/14999 | Loss: 0.0034 | Episodes: 29 | Win count: 1 | Win rate: 0.000 | ti
me: 18.56 minutes
Epoch: 004/14999 | Loss: 0.0485 | Episodes: 104 | Win count: 1 | Win rate: 0.000 | t
ime: 24.11 minutes
```

```
Epoch: 005/14999 | Loss: 0.0041 | Episodes: 106 | Win count: 1 | Win rate: 0.000 | t
ime: 29.78 minutes
Epoch: 006/14999 | Loss: 0.0135 | Episodes: 106 | Win count: 1 | Win rate: 0.000 | t
ime: 35.53 minutes
Epoch: 007/14999 | Loss: 0.0029 | Episodes: 107 | Win count: 1 | Win rate: 0.000 | t
ime: 41.35 minutes
Epoch: 008/14999 | Loss: 0.0016 | Episodes: 104 | Win count: 1 | Win rate: 0.000 | t
ime: 46.99 minutes
Epoch: 009/14999 | Loss: 0.0013 | Episodes: 104 | Win count: 1 | Win rate: 0.000 | t
ime: 52.66 minutes
Epoch: 010/14999 | Loss: 0.0028 | Episodes: 101 | Win count: 1 | Win rate: 0.000 | t
ime: 58.26 minutes
Epoch: 011/14999 | Loss: 0.0010 | Episodes: 104 | Win count: 1 | Win rate: 0.000 | t
ime: 63.95 minutes
Epoch: 012/14999 | Loss: 0.0026 | Episodes: 103 | Win count: 1 | Win rate: 0.000 | t
ime: 1.16 hours
Epoch: 013/14999 | Loss: 0.0022 | Episodes: 102 | Win count: 1 | Win rate: 0.000 | t
ime: 1.25 hours
Epoch: 014/14999 | Loss: 0.0014 | Episodes: 65 | Win count: 2 | Win rate: 0.000 | ti
me: 1.31 hours
Epoch: 015/14999 | Loss: 0.0021 | Episodes: 106 | Win count: 2 | Win rate: 0.000 | t
ime: 1.41 hours
Epoch: 016/14999 | Loss: 0.0015 | Episodes: 111 | Win count: 2 | Win rate: 0.000 | t
ime: 1.51 hours
Epoch: 017/14999 | Loss: 0.0023 | Episodes: 109 | Win count: 2 | Win rate: 0.000 | t
ime: 1.60 hours
Epoch: 018/14999 | Loss: 0.0015 | Episodes: 55 | Win count: 3 | Win rate: 0.000 | ti
me: 1.65 hours
Epoch: 019/14999 | Loss: 0.0102 | Episodes: 104 | Win count: 3 | Win rate: 0.000 | t
ime: 1.75 hours
Epoch: 020/14999 | Loss: 0.0221 | Episodes: 102 | Win count: 3 | Win rate: 0.000 | t
ime: 1.84 hours
Epoch: 021/14999 | Loss: 0.0357 | Episodes: 108 | Win count: 3 | Win rate: 0.000 | t
ime: 1.93 hours
Epoch: 022/14999 | Loss: 0.0013 | Episodes: 105 | Win count: 3 | Win rate: 0.000 | t
ime: 2.02 hours
Epoch: 023/14999 | Loss: 0.0014 | Episodes: 106 | Win count: 3 | Win rate: 0.000 | t
ime: 2.12 hours
Epoch: 024/14999 | Loss: 0.0032 | Episodes: 23 | Win count: 4 | Win rate: 0.167 | ti
me: 2.14 hours
Epoch: 025/14999 | Loss: 0.0055 | Episodes: 104 | Win count: 4 | Win rate: 0.167 | t
ime: 2.23 hours
Epoch: 026/14999 | Loss: 0.0012 | Episodes: 103 | Win count: 4 | Win rate: 0.167 | t
ime: 2.33 hours
Epoch: 027/14999 | Loss: 0.0012 | Episodes: 3 | Win count: 5 | Win rate: 0.167 | tim
e: 2.33 hours
Epoch: 028/14999 | Loss: 0.1357 | Episodes: 107 | Win count: 5 | Win rate: 0.167 | t
ime: 2.42 hours
Epoch: 029/14999 | Loss: 0.0017 | Episodes: 10 | Win count: 6 | Win rate: 0.208 | ti
me: 2.43 hours
Epoch: 030/14999 | Loss: 0.0429 | Episodes: 103 | Win count: 7 | Win rate: 0.250 | t
ime: 2.52 hours
Epoch: 031/14999 | Loss: 0.0033 | Episodes: 4 | Win count: 8 | Win rate: 0.292 | tim
e: 2.52 hours
Epoch: 032/14999 | Loss: 0.0034 | Episodes: 104 | Win count: 8 | Win rate: 0.292 | t
ime: 2.62 hours
Epoch: 033/14999 | Loss: 0.0022 | Episodes: 103 | Win count: 8 | Win rate: 0.292 | t
ime: 2.71 hours
Epoch: 034/14999 | Loss: 0.0020 | Episodes: 107 | Win count: 8 | Win rate: 0.292 | t
ime: 2.80 hours
Epoch: 035/14999 | Loss: 0.0016 | Episodes: 104 | Win count: 8 | Win rate: 0.292 | t
ime: 2.89 hours
Epoch: 036/14999 | Loss: 0.0005 | Episodes: 68 | Win count: 9 | Win rate: 0.333 | ti
me: 2.95 hours
Epoch: 037/14999 | Loss: 0.0018 | Episodes: 103 | Win count: 9 | Win rate: 0.333 | t
ime: 3.04 hours
Epoch: 038/14999 | Loss: 0.0010 | Episodes: 2 | Win count: 10 | Win rate: 0.333 | ti
me: 3.04 hours
Epoch: 039/14999 | Loss: 0.0016 | Episodes: 104 | Win count: 10 | Win rate: 0.333 |
```

```
time: 3.14 hours
Epoch: 040/14999 | Loss: 0.0017 | Episodes: 3 | Win count: 11 | Win rate: 0.375 | ti
me: 3.14 hours
Epoch: 041/14999 | Loss: 0.0023 | Episodes: 62 | Win count: 12 | Win rate: 0.417 | t
ime: 3.20 hours
Epoch: 042/14999 | Loss: 0.0149 | Episodes: 104 | Win count: 12 | Win rate: 0.375 |
time: 3.29 hours
Epoch: 043/14999 | Loss: 0.0029 | Episodes: 3 | Win count: 13 | Win rate: 0.417 | ti
me: 3.30 hours
Epoch: 044/14999 | Loss: 0.0050 | Episodes: 14 | Win count: 14 | Win rate: 0.458 | t
ime: 3.31 hours
Epoch: 045/14999 | Loss: 0.0072 | Episodes: 61 | Win count: 15 | Win rate: 0.500 | t
ime: 3.37 hours
Epoch: 046/14999 | Loss: 0.0024 | Episodes: 15 | Win count: 16 | Win rate: 0.542 | t
ime: 3.38 hours
Epoch: 047/14999 | Loss: 0.0109 | Episodes: 18 | Win count: 17 | Win rate: 0.583 | t
ime: 3.39 hours
Epoch: 048/14999 | Loss: 0.0019 | Episodes: 104 | Win count: 17 | Win rate: 0.542 |
time: 3.49 hours
Epoch: 049/14999 | Loss: 0.0039 | Episodes: 53 | Win count: 18 | Win rate: 0.583 | t
ime: 3.54 hours
Epoch: 050/14999 | Loss: 0.0025 | Episodes: 15 | Win count: 19 | Win rate: 0.625 | t
ime: 3.55 hours
Epoch: 051/14999 | Loss: 0.0395 | Episodes: 10 | Win count: 20 | Win rate: 0.625 | t
ime: 3.56 hours
Epoch: 052/14999 | Loss: 0.0013 | Episodes: 25 | Win count: 21 | Win rate: 0.667 | t
ime: 3.58 hours
Epoch: 053/14999 | Loss: 0.0015 | Episodes: 23 | Win count: 22 | Win rate: 0.667 | t
ime: 3.60 hours
Epoch: 054/14999 | Loss: 0.0046 | Episodes: 117 | Win count: 23 | Win rate: 0.667 |
time: 3.70 hours
Epoch: 055/14999 | Loss: 0.0018 | Episodes: 37 | Win count: 24 | Win rate: 0.667 | t
ime: 3.73 hours
Epoch: 056/14999 | Loss: 0.0008 | Episodes: 11 | Win count: 25 | Win rate: 0.708 | t
ime: 3.74 hours
Epoch: 057/14999 | Loss: 0.0005 | Episodes: 29 | Win count: 26 | Win rate: 0.750 | t
ime: 3.77 hours
Epoch: 058/14999 | Loss: 0.0013 | Episodes: 33 | Win count: 27 | Win rate: 0.792 | t
ime: 3.80 hours
Epoch: 059/14999 | Loss: 0.0011 | Episodes: 18 | Win count: 28 | Win rate: 0.833 | t
ime: 3.82 hours
```

Reinforcement Learning

MountainCar-v0

In [ ]:

```
class State:
  def __init__(self, position_tile, velocity_tile):
    self.position_tile = position_tile
    self.velocity_tile = velocity_tile

  def __eq__(self, other):
    if isinstance(other, State):
      return (self.position_tile == other.position_tile and
              self.velocity_tile == other.velocity_tile)
    else:
      return False

  def __hash__(self):
    return hash((self.position_tile, self.velocity_tile))
```

In [ ]:

```
ACTIONS_COUNT = 3
eps_greedy = 0
```

```
def find_best_action_quality(Q, state):
  best_action, best_q = None, None
  for action in range(ACTIONS_COUNT):
    cur_q = Q[(state, action)]
    if best_q is None or best_q < cur_q:
      best_action, best_q = action, cur_q
  return best_action, best_q


def choose_eps_greedy_action(Q, state):
  best_action, best_q = find_best_action_quality(Q, state)
  best_count = 0
  for action in range(ACTIONS_COUNT):
    if Q[(state, action)] == best_q:
        best_count += 1
  p = []
  for action in range(ACTIONS_COUNT):
    prob = eps_greedy / ACTIONS_COUNT
    if Q[(state, action)] == best_q:
      prob += (1 - eps_greedy) / best_count
    p.append(prob)
  return np.random.choice(ACTIONS_COUNT, 1, p=p)[0]


def map_observation_to_state(observation, position_grid, velocity_grid):
  return State(int(round(observation[0] / position_grid)),
               int(round(observation[1] / velocity_grid)))
```

In [ ]:

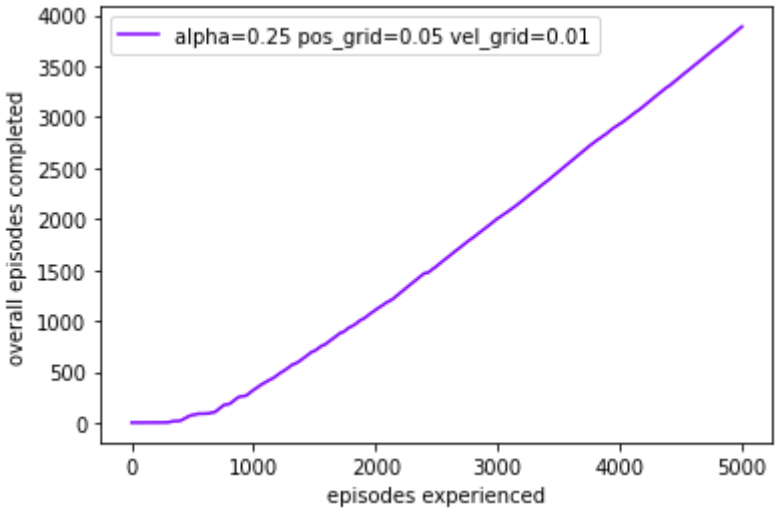
```
training_episodes = 5000
timesteps_limit = 200  # limit from OpenAI gym docs
gamma = 0.9

def evaluate_parameters(alpha, position_grid, velocity_grid):
  env = gym.make('MountainCar-v0')
  env.seed(0)
  np.random.seed(0)
  cumulative_completion = []
  completed = 0
  Q = defaultdict(lambda: 0.0)
  for episode in range(training_episodes):
    observation = env.reset()
    state = map_observation_to_state(
      observation, position_grid, velocity_grid)
    action = choose_eps_greedy_action(Q, state)
    for timestep in range(timesteps_limit):
      observation, reward, done, info = env.step(action)
      to_state = map_observation_to_state(
        observation, position_grid, velocity_grid)
      next_action = choose_eps_greedy_action(Q, to_state)
      Q[(state, action)] += alpha * (reward +
                                     gamma * Q[(to_state, action)] -
                                     Q[(state, action)])
      action, state = next_action, to_state
      if done:
        if timestep != timesteps_limit - 1:
          completed += 1
        cumulative_completion.append(completed)
        break
  env.close()
  return cumulative_completion
```

In [ ]:
```
def evaluate_and_plot_parameters(alpha, position_grid, velocity_grid, color):
  cumulative_completion = evaluate_parameters(alpha, position_grid, velocity_grid)
  title = f'alpha={alpha} pos_grid={position_grid} vel_grid={velocity_grid}'
  print(f'Evaluated {title}')
  line, = plt.plot(
      np.arange(1, training_episodes + 1),
      cumulative_completion,
      c=next(color),
      label=title)
  return line
```

In [ ]:
```
color = iter(cm.rainbow(np.linspace(0, 1, 5)))
handles = []

alpha, position_grid, velocity_grid = 0.25, 0.05, 0.01
handles.append(evaluate_and_plot_parameters(
  alpha, position_grid, velocity_grid, color
))

plt.xlabel('episodes experienced')
plt.ylabel('overall episodes completed')
plt.legend(handles=handles)
```

```
Evaluated alpha=0.25 pos_grid=0.05 vel_grid=0.01
```

Out[ ]:
```
<matplotlib.legend.Legend at 0x7f269f4c4b10>
```



Roulette

In [ ]:
```
class Model(object):
        def __init__(self, *args):
                if not args is None:
                        self.env = args[0]
                        self.Q = args[1]
                        self.alpha = args[2]
                        self.gamma = args[3]
                        self.epsilon = args[4]
                        self.n_episodes = args[5]
                        self.verbose = args[6]
                        self.record_training = args[7]
                        self.checkpoint = self.n_episodes * 0.1
                else:
                        print('Invalid arguments.')

        def eps_greedy(self, obs):
```

```
                if np.random.uniform() < self.epsilon:
                        return np.random.randint(self.env.action_space.n)
                else:
                        action_values = [self.Q[obs, a] for a in
                                                        range(self.env.action_space
                        greedy_idx = np.argwhere(action_values == np.max(action_valu
                        greedy_act_idx = np.random.choice(greedy_idx.flatten())
                        return greedy_act_idx

        def greedy_action(self, obs):
                action_values = [self.Q[obs, a] for a in
                                                range(self.env.action_space.n)]
                greedy_idx = np.argmax(action_values)
                return greedy_idx

        def train(self, idx=None, q=None):
                if self.record_training:
                        self.all_rewards = []

                for episode in range(self.n_episodes):
                        done = False
                        obs = self.env.reset()
                        if self.record_training:
                                episode_reward = 0
                        a = self.eps_greedy(obs)

                        while not done:
                                obs_prime, reward, done, info = self.env.step(a)
                                a_prime = self.eps_greedy(obs_prime)
                                self.Q[obs,a] += self.alpha * (reward + self.gamma*s

                                if self.record_training:
                                        episode_reward += reward
                                obs = obs_prime
                                a = a_prime


                        if self.record_training:
                                self.all_rewards.append(episode_reward)
                        if self.verbose and episode % self.checkpoint == 0:
                                if not idx is None:
                                        print(f'Agent: {idx} Episode: {episode}')
                                else:
                                        print(f'Episode: {episode}')
                if not q is None:
                        q.put(self)
                if not idx is None:
                        print(f'Agent: {idx} - Training complete.')
                else:
                        print('Training complete.')
```

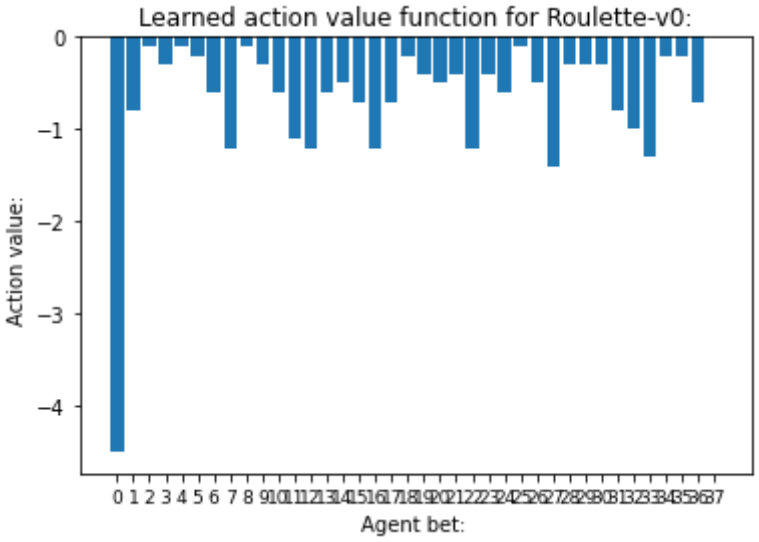
In [ ]:

```
# Initialize environment, hyperparameters and action value function.
gamma = 1
alpha = 0.1
epsilon = 0.1
n_epsiodes = 10000
env = gym.make('Roulette-v0')
Q = dict.fromkeys(product([0], range(38)), 0.0)

# Create and train agent.
agent = Model(env, Q, alpha, gamma, epsilon, n_epsiodes, True, False)
agent.train()
```

```
Episode: 0
Episode: 1000
Episode: 2000
Episode: 3000
Episode: 4000
Episode: 5000
Episode: 6000
Episode: 7000
Episode: 8000
Episode: 9000
Training complete.
```

In [ ]:

```
action_values = np.array([i for i in agent.Q.values()])
plt.bar(range(len(action_values)), action_values)
plt.xticks(range(len(action_values)))
plt.tick_params(axis='x', which='major', labelsize=9)
plt.xlabel('Agent bet: ')
plt.ylabel('Action value: ')
plt.title('Learned action value function for Roulette-v0:')
plt.show()
```



Deep Reinforcement Learning

MountainCar-v0

In [ ]:

```
class MountainCarTrain:
    def __init__(self,env):
        self.env=env
        self.gamma=0.99

        self.epsilon = 1
        self.epsilon_decay = 0.05

        self.epsilon_min=0.01


        self.learingRate=0.001

        self.replayBuffer=deque(maxlen=20000)
        self.trainNetwork=self.createNetwork()

        self.episodeNum=400
```

```
        self.iterationNum=201 #max is 200

        self.numPickFromBuffer=32

        self.targetNetwork=self.createNetwork()

        self.targetNetwork.set_weights(self.trainNetwork.get_weights())

    def createNetwork(self):
        model = models.Sequential()
        state_shape = self.env.observation_space.shape

        model.add(layers.Dense(24, activation='relu', input_shape=state_shape))
        model.add(layers.Dense(48, activation='relu'))
        model.add(layers.Dense(self.env.action_space.n,activation='linear'))
        # model.compile(optimizer=optimizers.RMSprop(lr=self.learingRate), loss=loss
        model.compile(loss='mse', optimizer=Adam(learning_rate=self.learingRate))
        return model

    def getBestAction(self,state):

        self.epsilon = max(self.epsilon_min, self.epsilon)

        if np.random.rand(1) < self.epsilon:
            action = np.random.randint(0, 3)
        else:
            action=np.argmax(self.trainNetwork.predict(state)[0])

        return action



    def trainFromBuffer_Boost(self):
        if len(self.replayBuffer) < self.numPickFromBuffer:
            return
        samples = random.sample(self.replayBuffer,self.numPickFromBuffer)
        npsamples = np.array(samples)
        states_temp, actions_temp, rewards_temp, newstates_temp, dones_temp = np.hsp
        states = np.concatenate((np.squeeze(states_temp[:])), axis = 0)
        rewards = rewards_temp.reshape(self.numPickFromBuffer,).astype(float)
        targets = self.trainNetwork.predict(states)
        newstates = np.concatenate(np.concatenate(newstates_temp))
        dones = np.concatenate(dones_temp).astype(bool)
        notdones = ~dones
        notdones = notdones.astype(float)
        dones = dones.astype(float)
        Q_futures = self.targetNetwork.predict(newstates).max(axis = 1)
        targets[(np.arange(self.numPickFromBuffer), actions_temp.reshape(self.numPic
        self.trainNetwork.fit(states, targets, epochs=1, verbose=0)



    def trainFromBuffer(self):
        if len(self.replayBuffer) < self.numPickFromBuffer:
            return

        samples = random.sample(self.replayBuffer,self.numPickFromBuffer)

        states = []
        newStates=[]
        for sample in samples:
            state, action, reward, new_state, done = sample
            states.append(state)
            newStates.append(new_state)
```

```
        newArray = np.array(states)
        states = newArray.reshape(self.numPickFromBuffer, 2)

        newArray2 = np.array(newStates)
        newStates = newArray2.reshape(self.numPickFromBuffer, 2)

        targets = self.trainNetwork.predict(states)
        new_state_targets=self.targetNetwork.predict(newStates)

        i=0
        for sample in samples:
            state, action, reward, new_state, done = sample
            target = targets[i]
            if done:
                target[action] = reward
            else:
                Q_future = max(new_state_targets[i])
                target[action] = reward + Q_future * self.gamma
            i+=1

        self.trainNetwork.fit(states, targets, epochs=1, verbose=0)


    def orginalTry(self,currentState,eps):
        rewardSum = 0
        max_position=-99

        for i in range(self.iterationNum):
            bestAction = self.getBestAction(currentState)

            new_state, reward, done, _ = env.step(bestAction)

            new_state = new_state.reshape(1, 2)

            # # Keep track of max position
            if new_state[0][0] > max_position:
                max_position = new_state[0][0]


            # # Adjust reward for task completion
            if new_state[0][0] >= 0.5:
                reward += 10

            self.replayBuffer.append([currentState, bestAction, reward, new_state, d

            #Or you can use self.trainFromBuffer_Boost(), it is a matrix wise versio
            self.trainFromBuffer()

            rewardSum += reward

            currentState = new_state

            if done:
                break

        if i >= 199:
            print("Failed to finish task in epsoide {}".format(eps))
        else:
            print("Success in epsoide {}, used {} iterations!".format(eps, i))
            self.trainNetwork.save('./trainNetworkInEPS{}.h5'.format(eps))

        #Sync
        self.targetNetwork.set_weights(self.trainNetwork.get_weights())
```

```
        print("now epsilon is {}, the reward is {} maxPosition is {}".format(max(sel
        self.epsilon -= self.epsilon_decay

    def start(self):
        for eps in range(self.episodeNum):
            currentState=env.reset().reshape(1,2)
            self.orginalTry(currentState, eps)
```

In [ ]:

```
env = gym.make('MountainCar-v0')
dqn=MountainCarTrain(env=env)
dqn.start()
```

```
Failed to finish task in epsoide 0
now epsilon is 1, the reward is -200.0 maxPosition is -0.3524132933850725
Failed to finish task in epsoide 1
now epsilon is 0.95, the reward is -200.0 maxPosition is -0.3070783992907518
Failed to finish task in epsoide 2
now epsilon is 0.8999999999999999, the reward is -200.0 maxPosition is -0.4299435894
539687
Failed to finish task in epsoide 3
now epsilon is 0.8499999999999999, the reward is -200.0 maxPosition is -0.4344620785
8135955
Failed to finish task in epsoide 4
now epsilon is 0.7999999999999998, the reward is -200.0 maxPosition is -0.2803382396
02065
Failed to finish task in epsoide 5
now epsilon is 0.7499999999999998, the reward is -200.0 maxPosition is -0.3681538690
378387
Failed to finish task in epsoide 6
now epsilon is 0.6999999999999997, the reward is -200.0 maxPosition is -0.3626199720
2587165
Failed to finish task in epsoide 7
now epsilon is 0.6499999999999997, the reward is -200.0 maxPosition is -0.2033583982
262082
Failed to finish task in epsoide 8
now epsilon is 0.5999999999999996, the reward is -200.0 maxPosition is -0.1385556902
3469297
Failed to finish task in epsoide 9
now epsilon is 0.5499999999999996, the reward is -200.0 maxPosition is -0.1484217832
4856836
Failed to finish task in epsoide 10
now epsilon is 0.4999999999999996, the reward is -200.0 maxPosition is -0.2793319258
043251
Failed to finish task in epsoide 11
now epsilon is 0.4499999999999996, the reward is -200.0 maxPosition is -0.3795558553
134636
Failed to finish task in epsoide 12
now epsilon is 0.39999999999999963, the reward is -200.0 maxPosition is -0.152149683
59420947
Failed to finish task in epsoide 13
now epsilon is 0.34999999999999964, the reward is -200.0 maxPosition is -0.234883204
7537745
Failed to finish task in epsoide 14
now epsilon is 0.29999999999999966, the reward is -200.0 maxPosition is -0.149767399
2059681
Failed to finish task in epsoide 15
now epsilon is 0.24999999999999967, the reward is -200.0 maxPosition is -0.084132581
19435982
Failed to finish task in epsoide 16
now epsilon is 0.19999999999999968, the reward is -200.0 maxPosition is -0.278863784
7690169
Failed to finish task in epsoide 17
now epsilon is 0.1499999999999997, the reward is -200.0 maxPosition is -0.1754113932
7945746
Failed to finish task in epsoide 18
now epsilon is 0.09999999999999969, the reward is -200.0 maxPosition is -0.238859330
```

```
16467759
Failed to finish task in epsoide 19
now epsilon is 0.049999999999999684, the reward is -200.0 maxPosition is -0.11590947
866633974
Failed to finish task in epsoide 20
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.12057596031550417
Success in epsoide 21, used 185 iterations!
now epsilon is 0.01, the reward is -176.0 maxPosition is 0.5294146062416286
Failed to finish task in epsoide 22
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.18958001342887487
Failed to finish task in epsoide 23
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.07687881686420892
Failed to finish task in epsoide 24
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.20245064840757712
Failed to finish task in epsoide 25
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.07706479055932372
Failed to finish task in epsoide 26
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.10456900634389282
Failed to finish task in epsoide 27
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.02419682800432305
Failed to finish task in epsoide 28
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.3085194327013536
Failed to finish task in epsoide 29
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.1193437903467623
Failed to finish task in epsoide 30
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.13010469556431345
Failed to finish task in epsoide 31
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.253077924123337
Failed to finish task in epsoide 32
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.04981205956334721
Failed to finish task in epsoide 33
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.09580553163391546
Failed to finish task in epsoide 34
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.004608990170606679
Failed to finish task in epsoide 35
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.2760154039240527
Failed to finish task in epsoide 36
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.19128354963158592
Failed to finish task in epsoide 37
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.03475348570020361
Failed to finish task in epsoide 38
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.22268068276799002
Failed to finish task in epsoide 39
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.2033148717134867
Failed to finish task in epsoide 40
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.2145653751979175
Failed to finish task in epsoide 41
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.09969594676308417
Failed to finish task in epsoide 42
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.04221954398141184
Failed to finish task in epsoide 43
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.23829267350123073
Failed to finish task in epsoide 44
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.22811383970325916
Failed to finish task in epsoide 45
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.019928887042185605
Failed to finish task in epsoide 46
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.16465049047824673
Failed to finish task in epsoide 47
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.23211727866456927
Success in epsoide 48, used 115 iterations!
now epsilon is 0.01, the reward is -106.0 maxPosition is 0.5146450313176049
Failed to finish task in epsoide 49
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.03757323940397939
Failed to finish task in epsoide 50
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.1327045195014213
Failed to finish task in epsoide 51
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.17859691787926024
Failed to finish task in epsoide 52
```

```
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.28437722651457875
Failed to finish task in epsoide 53
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.31784345259358654
Success in epsoide 54, used 166 iterations!
now epsilon is 0.01, the reward is -157.0 maxPosition is 0.5187573452834087
Failed to finish task in epsoide 55
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.2495120525554502
Failed to finish task in epsoide 56
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.026103184552185105
Failed to finish task in epsoide 57
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.1928315780895043
Failed to finish task in epsoide 58
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.16206150286733176
Failed to finish task in epsoide 59
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.12057070647127391
Success in epsoide 60, used 98 iterations!
now epsilon is 0.01, the reward is -89.0 maxPosition is 0.5154675016263012
Failed to finish task in epsoide 61
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.3266907765543536
Success in epsoide 62, used 115 iterations!
now epsilon is 0.01, the reward is -106.0 maxPosition is 0.5093562372516088
Success in epsoide 63, used 102 iterations!
now epsilon is 0.01, the reward is -93.0 maxPosition is 0.5032665865268193
Failed to finish task in epsoide 64
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.03158751851330652
Failed to finish task in epsoide 65
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.2424638410147817
Failed to finish task in epsoide 66
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.1019430074053843
Failed to finish task in epsoide 67
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.14255415900949733
Failed to finish task in epsoide 68
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.05053562716336016
Failed to finish task in epsoide 69
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.14940601513888294
Success in epsoide 70, used 158 iterations!
now epsilon is 0.01, the reward is -149.0 maxPosition is 0.5211466536320155
Failed to finish task in epsoide 71
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.22828822537240048
Failed to finish task in epsoide 72
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.28256147766800965
Failed to finish task in epsoide 73
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.3577043644844212
Failed to finish task in epsoide 74
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.13450531822988185
Success in epsoide 75, used 174 iterations!
now epsilon is 0.01, the reward is -165.0 maxPosition is 0.5059110852620284
Failed to finish task in epsoide 76
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.138928734339838
Failed to finish task in epsoide 77
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.27758137421887413
Failed to finish task in epsoide 78
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.25317017857459756
Failed to finish task in epsoide 79
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.2258238267981899
Failed to finish task in epsoide 80
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.3130529841388198
Failed to finish task in epsoide 81
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.08706571126531527
Failed to finish task in epsoide 82
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.1462486634505181
Failed to finish task in epsoide 83
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.34276160870772293
Failed to finish task in epsoide 84
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.0870642490145532
Failed to finish task in epsoide 85
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.01879483725826105
Failed to finish task in epsoide 86
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.3674493707721311
```

```
Failed to finish task in epsoide 87
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.28638191156605713
Success in epsoide 88, used 159 iterations!
now epsilon is 0.01, the reward is -150.0 maxPosition is 0.5076502699253835
Failed to finish task in epsoide 89
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.35281639577528084
Failed to finish task in epsoide 90
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.2932727044705906
Failed to finish task in epsoide 91
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.06956679438850356
Failed to finish task in epsoide 92
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.26076986791356377
Failed to finish task in epsoide 93
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.2408647638434624
Failed to finish task in epsoide 94
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.08280775649750768
Failed to finish task in epsoide 95
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.16373455704236908
Failed to finish task in epsoide 96
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.1484855716729034
Failed to finish task in epsoide 97
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.4066075962847191
Failed to finish task in epsoide 98
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.03091052433703733
Failed to finish task in epsoide 99
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.2749867307157541
Failed to finish task in epsoide 100
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.0013214572967789905
Failed to finish task in epsoide 101
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.2171181534465875
Failed to finish task in epsoide 102
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.1241058264474776
Failed to finish task in epsoide 103
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.010188534873267888
Failed to finish task in epsoide 104
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.0021787170758499654
Failed to finish task in epsoide 105
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.4171489240764023
Failed to finish task in epsoide 106
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.40799331971256275
Failed to finish task in epsoide 107
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.10868328455358472
Failed to finish task in epsoide 108
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.31992749019392
Success in epsoide 109, used 152 iterations!
now epsilon is 0.01, the reward is -143.0 maxPosition is 0.5221052039242029
Failed to finish task in epsoide 110
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.10003874232306688
Failed to finish task in epsoide 111
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.02444470209764
Failed to finish task in epsoide 112
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.15709576188779367
Failed to finish task in epsoide 113
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.03843210057705962
Failed to finish task in epsoide 114
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.04351552286369948
Failed to finish task in epsoide 115
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.1468716207805081
Failed to finish task in epsoide 116
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.07642418458642222
Failed to finish task in epsoide 117
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.36747026615683426
Failed to finish task in epsoide 118
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.15824449466394083
Failed to finish task in epsoide 119
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.07298235077250485
Success in epsoide 120, used 163 iterations!
now epsilon is 0.01, the reward is -154.0 maxPosition is 0.5250838395911595
Failed to finish task in epsoide 121
```

```
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.00787213243487874
Failed to finish task in epsoide 122
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.0854474665777644
Failed to finish task in epsoide 123
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.1417763280201628
Success in epsoide 124, used 132 iterations!
now epsilon is 0.01, the reward is -123.0 maxPosition is 0.5212711802880642
Failed to finish task in epsoide 125
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.27775662796627576
Failed to finish task in epsoide 126
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.19929083510645101
Success in epsoide 127, used 195 iterations!
now epsilon is 0.01, the reward is -186.0 maxPosition is 0.5177758039357164
Failed to finish task in epsoide 128
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.07281634191825924
Failed to finish task in epsoide 129
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.23514054916909405
Failed to finish task in epsoide 130
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.057916105681928925
Success in epsoide 131, used 186 iterations!
now epsilon is 0.01, the reward is -177.0 maxPosition is 0.5191966956131646
Failed to finish task in epsoide 132
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.2431024558792686
Failed to finish task in epsoide 133
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.33212900649025073
Failed to finish task in epsoide 134
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.2896495091176193
Failed to finish task in epsoide 135
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.23403085407605875
Failed to finish task in epsoide 136
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.19508774569403695
Failed to finish task in epsoide 137
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.10848715865766262
Failed to finish task in epsoide 138
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.3807164705646364
Failed to finish task in epsoide 139
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.53637911462134
Failed to finish task in epsoide 140
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.5067311088727282
Failed to finish task in epsoide 141
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.04346658918499973
Failed to finish task in epsoide 142
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.28551421884480604
Failed to finish task in epsoide 143
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.16662953999750757
Failed to finish task in epsoide 144
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.1258727848972562
Success in epsoide 145, used 186 iterations!
now epsilon is 0.01, the reward is -177.0 maxPosition is 0.5071286824350305
Failed to finish task in epsoide 146
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.3527220887394995
Success in epsoide 147, used 196 iterations!
now epsilon is 0.01, the reward is -187.0 maxPosition is 0.5025737952388928
Failed to finish task in epsoide 148
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.34994944722049426
Failed to finish task in epsoide 149
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.1679413263924455
Failed to finish task in epsoide 150
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.12593262817692463
Failed to finish task in epsoide 151
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.5081660470660619
Failed to finish task in epsoide 152
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.40813675021640883
Success in epsoide 153, used 191 iterations!
now epsilon is 0.01, the reward is -182.0 maxPosition is 0.5001177961030273
Success in epsoide 154, used 99 iterations!
now epsilon is 0.01, the reward is -90.0 maxPosition is 0.5158885869590966
Failed to finish task in epsoide 155
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.08934228711530823
```

```
Failed to finish task in epsoide 156
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.12987132939915952
Failed to finish task in epsoide 157
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.3635141594455871
Failed to finish task in epsoide 158
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.015729962296154067
Success in epsoide 159, used 182 iterations!
now epsilon is 0.01, the reward is -173.0 maxPosition is 0.5245149251599214
Failed to finish task in epsoide 160
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.13743351588489944
Failed to finish task in epsoide 161
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.32877994226174195
Failed to finish task in epsoide 162
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.39322571768888115
Failed to finish task in epsoide 163
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.1328073667469478
Failed to finish task in epsoide 164
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.007507882628279195
Failed to finish task in epsoide 165
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.34710568506345274
Failed to finish task in epsoide 166
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.050038344962540775
Failed to finish task in epsoide 167
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.5323028992044582
Failed to finish task in epsoide 168
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.1733235302403134
Failed to finish task in epsoide 169
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.2723347978802802
Success in epsoide 170, used 183 iterations!
now epsilon is 0.01, the reward is -174.0 maxPosition is 0.5085949577799603
Success in epsoide 171, used 178 iterations!
now epsilon is 0.01, the reward is -169.0 maxPosition is 0.5156499066168116
Failed to finish task in epsoide 172
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.29331676393455547
Failed to finish task in epsoide 173
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.03734431939163221
Success in epsoide 174, used 195 iterations!
now epsilon is 0.01, the reward is -186.0 maxPosition is 0.5173555183544168
Success in epsoide 175, used 127 iterations!
now epsilon is 0.01, the reward is -118.0 maxPosition is 0.5076784760942776
Failed to finish task in epsoide 176
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.46917394976471816
Success in epsoide 177, used 179 iterations!
now epsilon is 0.01, the reward is -170.0 maxPosition is 0.5200948158319417
Failed to finish task in epsoide 178
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.26599991121625777
Failed to finish task in epsoide 179
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.40696723025268444
Success in epsoide 180, used 176 iterations!
now epsilon is 0.01, the reward is -167.0 maxPosition is 0.5017570810916677
Success in epsoide 181, used 184 iterations!
now epsilon is 0.01, the reward is -175.0 maxPosition is 0.5074174909892076
Success in epsoide 182, used 185 iterations!
now epsilon is 0.01, the reward is -176.0 maxPosition is 0.5192736070568075
Success in epsoide 183, used 195 iterations!
now epsilon is 0.01, the reward is -186.0 maxPosition is 0.5061002953511963
Success in epsoide 184, used 168 iterations!
now epsilon is 0.01, the reward is -159.0 maxPosition is 0.5126335280808539
Success in epsoide 185, used 135 iterations!
now epsilon is 0.01, the reward is -126.0 maxPosition is 0.5222766334213044
Failed to finish task in epsoide 186
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.496170714535724
Success in epsoide 187, used 188 iterations!
now epsilon is 0.01, the reward is -179.0 maxPosition is 0.5038491073612663
Success in epsoide 188, used 189 iterations!
now epsilon is 0.01, the reward is -180.0 maxPosition is 0.5289076277013133
Success in epsoide 189, used 173 iterations!
now epsilon is 0.01, the reward is -164.0 maxPosition is 0.5132240455375635
Success in epsoide 190, used 132 iterations!
```

```
now epsilon is 0.01, the reward is -123.0 maxPosition is 0.5007067825792395
Success in epsoide 191, used 137 iterations!
now epsilon is 0.01, the reward is -128.0 maxPosition is 0.5036285384244206
Success in epsoide 192, used 146 iterations!
now epsilon is 0.01, the reward is -137.0 maxPosition is 0.5251084107491971
Failed to finish task in epsoide 193
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.25280145222186906
Success in epsoide 194, used 184 iterations!
now epsilon is 0.01, the reward is -175.0 maxPosition is 0.5168520693424925
Success in epsoide 195, used 102 iterations!
now epsilon is 0.01, the reward is -93.0 maxPosition is 0.5147889152332306
Failed to finish task in epsoide 196
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.3044876615782575
Success in epsoide 197, used 176 iterations!
now epsilon is 0.01, the reward is -167.0 maxPosition is 0.5142995709631081
Failed to finish task in epsoide 198
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.20051720120671024
Failed to finish task in epsoide 199
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.30504320056441964
Success in epsoide 200, used 116 iterations!
now epsilon is 0.01, the reward is -107.0 maxPosition is 0.5154271293080845
Failed to finish task in epsoide 201
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.2734085904363 58
Success in epsoide 202, used 128 iterations!
now epsilon is 0.01, the reward is -119.0 maxPosition is 0.5096890039957227
Failed to finish task in epsoide 203
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.11326490299649283
Failed to finish task in epsoide 204
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.2833137043151989
Success in epsoide 205, used 121 iterations!
now epsilon is 0.01, the reward is -112.0 maxPosition is 0.5061840405109604
Failed to finish task in epsoide 206
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.24257898087796734
Failed to finish task in epsoide 207
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.05717395325635527
Success in epsoide 208, used 166 iterations!
now epsilon is 0.01, the reward is -157.0 maxPosition is 0.508812141263922
Failed to finish task in epsoide 209
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.12176754594095415
Failed to finish task in epsoide 210
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.18307186721219373
Success in epsoide 211, used 180 iterations!
now epsilon is 0.01, the reward is -171.0 maxPosition is 0.5088893895937298
Failed to finish task in epsoide 212
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.13660276698026647
Failed to finish task in epsoide 213
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.24082172003829
Failed to finish task in epsoide 214
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.12325192443901363
Failed to finish task in epsoide 215
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.29235619311248306
Failed to finish task in epsoide 216
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.18075233896314713
Failed to finish task in epsoide 217
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.11972948014929852
Failed to finish task in epsoide 218
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.0016394815316916815
Success in epsoide 219, used 179 iterations!
now epsilon is 0.01, the reward is -170.0 maxPosition is 0.5156499066168116
Failed to finish task in epsoide 220
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.059449682514593136
Failed to finish task in epsoide 221
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.03725454636448963
Failed to finish task in epsoide 222
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.2522611171096353
Failed to finish task in epsoide 223
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.04933299303819067
Failed to finish task in epsoide 224
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.00456951959023612
```

```
Success in epsoide 225, used 160 iterations!
now epsilon is 0.01, the reward is -151.0 maxPosition is 0.510738735443973
Success in epsoide 226, used 103 iterations!
now epsilon is 0.01, the reward is -94.0 maxPosition is 0.5055861836853976
Success in epsoide 227, used 181 iterations!
now epsilon is 0.01, the reward is -172.0 maxPosition is 0.5125805611439715
Success in epsoide 228, used 183 iterations!
now epsilon is 0.01, the reward is -174.0 maxPosition is 0.5051443499346304
Success in epsoide 229, used 155 iterations!
now epsilon is 0.01, the reward is -146.0 maxPosition is 0.5104080635903744
Success in epsoide 230, used 151 iterations!
now epsilon is 0.01, the reward is -142.0 maxPosition is 0.5205603734740933
Failed to finish task in epsoide 231
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.015384440032770482
Success in epsoide 232, used 150 iterations!
now epsilon is 0.01, the reward is -141.0 maxPosition is 0.5225951874497831
Success in epsoide 233, used 169 iterations!
now epsilon is 0.01, the reward is -160.0 maxPosition is 0.5326581113024246
Failed to finish task in epsoide 234
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.22360406767354488
Success in epsoide 235, used 170 iterations!
now epsilon is 0.01, the reward is -161.0 maxPosition is 0.523663181689715
Failed to finish task in epsoide 236
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.4395651150509545
Success in epsoide 237, used 152 iterations!
now epsilon is 0.01, the reward is -143.0 maxPosition is 0.5314932056475683
Failed to finish task in epsoide 238
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.35087120088274204
Success in epsoide 239, used 167 iterations!
now epsilon is 0.01, the reward is -158.0 maxPosition is 0.501399043641431
Failed to finish task in epsoide 240
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.402662975951206
Success in epsoide 241, used 170 iterations!
now epsilon is 0.01, the reward is -161.0 maxPosition is 0.512541364035775
Failed to finish task in epsoide 242
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.08058398722110718
Failed to finish task in epsoide 243
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.4572944086125372
Success in epsoide 244, used 140 iterations!
now epsilon is 0.01, the reward is -131.0 maxPosition is 0.5052800100824854
Failed to finish task in epsoide 245
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.38919377315275316
Failed to finish task in epsoide 246
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.26831229200639123
Success in epsoide 247, used 176 iterations!
now epsilon is 0.01, the reward is -167.0 maxPosition is 0.5059158386879964
Failed to finish task in epsoide 248
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.424643607585222
Success in epsoide 249, used 168 iterations!
now epsilon is 0.01, the reward is -159.0 maxPosition is 0.503106542999169
Success in epsoide 250, used 171 iterations!
now epsilon is 0.01, the reward is -162.0 maxPosition is 0.5129109719935041
Failed to finish task in epsoide 251
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.13167255120063995
Failed to finish task in epsoide 252
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.18216387262067923
Success in epsoide 253, used 163 iterations!
now epsilon is 0.01, the reward is -154.0 maxPosition is 0.5022357688707656
Failed to finish task in epsoide 254
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.16014519112826636
Failed to finish task in epsoide 255
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.16115114774040248
Success in epsoide 256, used 161 iterations!
now epsilon is 0.01, the reward is -152.0 maxPosition is 0.5011844219837774
Success in epsoide 257, used 193 iterations!
now epsilon is 0.01, the reward is -184.0 maxPosition is 0.5045368641193779
Failed to finish task in epsoide 258
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.460610042494133
Failed to finish task in epsoide 259
```

```
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.2434202588483171
Success in epsoide 260, used 119 iterations!
now epsilon is 0.01, the reward is -110.0 maxPosition is 0.5022682003904714
Failed to finish task in epsoide 261
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.4963203243129449
Success in epsoide 262, used 84 iterations!
now epsilon is 0.01, the reward is -75.0 maxPosition is 0.5068621771946921
Success in epsoide 263, used 162 iterations!
now epsilon is 0.01, the reward is -153.0 maxPosition is 0.510454900543527
Success in epsoide 264, used 87 iterations!
now epsilon is 0.01, the reward is -78.0 maxPosition is 0.514679671207304
Success in epsoide 265, used 153 iterations!
now epsilon is 0.01, the reward is -144.0 maxPosition is 0.5220490622322357
Success in epsoide 266, used 154 iterations!
now epsilon is 0.01, the reward is -145.0 maxPosition is 0.5122627620704988
Success in epsoide 267, used 155 iterations!
now epsilon is 0.01, the reward is -146.0 maxPosition is 0.5015085364567435
Success in epsoide 268, used 154 iterations!
now epsilon is 0.01, the reward is -145.0 maxPosition is 0.5028316823875842
Success in epsoide 269, used 151 iterations!
now epsilon is 0.01, the reward is -142.0 maxPosition is 0.5368577983788596
Success in epsoide 270, used 85 iterations!
now epsilon is 0.01, the reward is -76.0 maxPosition is 0.5018168325185327
Success in epsoide 271, used 94 iterations!
now epsilon is 0.01, the reward is -85.0 maxPosition is 0.5011861669828531
Success in epsoide 272, used 91 iterations!
now epsilon is 0.01, the reward is -82.0 maxPosition is 0.5004692093438424
Success in epsoide 273, used 90 iterations!
now epsilon is 0.01, the reward is -81.0 maxPosition is 0.5190579681491923
Success in epsoide 274, used 158 iterations!
now epsilon is 0.01, the reward is -149.0 maxPosition is 0.5396609947637298
Success in epsoide 275, used 175 iterations!
now epsilon is 0.01, the reward is -166.0 maxPosition is 0.503780048166395
Failed to finish task in epsoide 276
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.11993668135289244
Success in epsoide 277, used 158 iterations!
now epsilon is 0.01, the reward is -149.0 maxPosition is 0.5368577983788596
Failed to finish task in epsoide 278
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.4282744910229501
Success in epsoide 279, used 168 iterations!
now epsilon is 0.01, the reward is -159.0 maxPosition is 0.5368577983788596
Failed to finish task in epsoide 280
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.05490740260597486
Success in epsoide 281, used 172 iterations!
now epsilon is 0.01, the reward is -163.0 maxPosition is 0.5368577983788596
Success in epsoide 282, used 183 iterations!
now epsilon is 0.01, the reward is -174.0 maxPosition is 0.5368577983788596
Failed to finish task in epsoide 283
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.12029166277232908
Success in epsoide 284, used 163 iterations!
now epsilon is 0.01, the reward is -154.0 maxPosition is 0.5368577983788596
Success in epsoide 285, used 123 iterations!
now epsilon is 0.01, the reward is -114.0 maxPosition is 0.5104924964055697
Success in epsoide 286, used 90 iterations!
now epsilon is 0.01, the reward is -81.0 maxPosition is 0.5047868193997761
Success in epsoide 287, used 84 iterations!
now epsilon is 0.01, the reward is -75.0 maxPosition is 0.5180044005560268
Success in epsoide 288, used 155 iterations!
now epsilon is 0.01, the reward is -146.0 maxPosition is 0.5368577983788596
Success in epsoide 289, used 158 iterations!
now epsilon is 0.01, the reward is -149.0 maxPosition is 0.5066004166176113
Success in epsoide 290, used 147 iterations!
now epsilon is 0.01, the reward is -138.0 maxPosition is 0.5043155639159533
Failed to finish task in epsoide 291
now epsilon is 0.01, the reward is -200.0 maxPosition is 0.2869967940337662
Success in epsoide 292, used 100 iterations!
now epsilon is 0.01, the reward is -91.0 maxPosition is 0.5037177311982022
Success in epsoide 293, used 131 iterations!
now epsilon is 0.01, the reward is -122.0 maxPosition is 0.5153602627291055
```

```
Success in epsoide 294, used 177 iterations!
now epsilon is 0.01, the reward is -168.0 maxPosition is 0.5178236928840169
Success in epsoide 295, used 144 iterations!
now epsilon is 0.01, the reward is -135.0 maxPosition is 0.5162936237262619
Success in epsoide 296, used 155 iterations!
now epsilon is 0.01, the reward is -146.0 maxPosition is 0.5368577983788596
Success in epsoide 297, used 92 iterations!
now epsilon is 0.01, the reward is -83.0 maxPosition is 0.5109950215918876
Success in epsoide 298, used 150 iterations!
now epsilon is 0.01, the reward is -141.0 maxPosition is 0.502802491577734
Success in epsoide 299, used 159 iterations!
now epsilon is 0.01, the reward is -150.0 maxPosition is 0.5126748855866896
Success in epsoide 300, used 90 iterations!
now epsilon is 0.01, the reward is -81.0 maxPosition is 0.5010716742088425
Success in epsoide 301, used 138 iterations!
now epsilon is 0.01, the reward is -129.0 maxPosition is 0.5368577983788596
Success in epsoide 302, used 164 iterations!
now epsilon is 0.01, the reward is -155.0 maxPosition is 0.5218396476447826
Success in epsoide 303, used 170 iterations!
now epsilon is 0.01, the reward is -161.0 maxPosition is 0.5368577983788596
Success in epsoide 304, used 157 iterations!
now epsilon is 0.01, the reward is -148.0 maxPosition is 0.5307989070535666
Success in epsoide 305, used 146 iterations!
now epsilon is 0.01, the reward is -137.0 maxPosition is 0.505662886654018
Success in epsoide 306, used 137 iterations!
now epsilon is 0.01, the reward is -128.0 maxPosition is 0.5368577983788596
Success in epsoide 307, used 156 iterations!
now epsilon is 0.01, the reward is -147.0 maxPosition is 0.5368577983788596
Success in epsoide 308, used 141 iterations!
now epsilon is 0.01, the reward is -132.0 maxPosition is 0.5368577983788596
Success in epsoide 309, used 149 iterations!
now epsilon is 0.01, the reward is -140.0 maxPosition is 0.5368577983788596
Success in epsoide 310, used 144 iterations!
now epsilon is 0.01, the reward is -135.0 maxPosition is 0.5055352839757251
Success in epsoide 311, used 154 iterations!
now epsilon is 0.01, the reward is -145.0 maxPosition is 0.5368577983788596
Failed to finish task in epsoide 312
now epsilon is 0.01, the reward is -200.0 maxPosition is -0.1482236132131905
Success in epsoide 313, used 154 iterations!
now epsilon is 0.01, the reward is -145.0 maxPosition is 0.5368577983788596
Success in epsoide 314, used 134 iterations!
now epsilon is 0.01, the reward is -125.0 maxPosition is 0.5368577983788596
Success in epsoide 315, used 149 iterations!
now epsilon is 0.01, the reward is -140.0 maxPosition is 0.5368577983788596
Success in epsoide 316, used 149 iterations!
now epsilon is 0.01, the reward is -140.0 maxPosition is 0.5168264933681779
Success in epsoide 317, used 150 iterations!
now epsilon is 0.01, the reward is -141.0 maxPosition is 0.5164555154369531
Success in epsoide 318, used 162 iterations!
now epsilon is 0.01, the reward is -153.0 maxPosition is 0.5368577983788596
Success in epsoide 319, used 167 iterations!
now epsilon is 0.01, the reward is -158.0 maxPosition is 0.5410568598079735
Success in epsoide 320, used 138 iterations!
now epsilon is 0.01, the reward is -129.0 maxPosition is 0.5344569582148987
Success in epsoide 321, used 128 iterations!
now epsilon is 0.01, the reward is -119.0 maxPosition is 0.5214169148439793
Success in epsoide 322, used 89 iterations!
now epsilon is 0.01, the reward is -80.0 maxPosition is 0.5150715461162264
Success in epsoide 323, used 87 iterations!
now epsilon is 0.01, the reward is -78.0 maxPosition is 0.5128036159395065
Success in epsoide 324, used 133 iterations!
now epsilon is 0.01, the reward is -124.0 maxPosition is 0.5368577983788596
Success in epsoide 325, used 152 iterations!
now epsilon is 0.01, the reward is -143.0 maxPosition is 0.5368577983788596
Success in epsoide 326, used 111 iterations!
now epsilon is 0.01, the reward is -102.0 maxPosition is 0.5063570394381067
Success in epsoide 327, used 112 iterations!
now epsilon is 0.01, the reward is -103.0 maxPosition is 0.5383601931733877
Success in epsoide 328, used 121 iterations!
```

```
now epsilon is 0.01, the reward is -112.0 maxPosition is 0.5103238791603762
Success in epsoide 329, used 111 iterations!
now epsilon is 0.01, the reward is -102.0 maxPosition is 0.505103877914227
Success in epsoide 330, used 149 iterations!
now epsilon is 0.01, the reward is -140.0 maxPosition is 0.5368577983788596
Success in epsoide 331, used 135 iterations!
now epsilon is 0.01, the reward is -126.0 maxPosition is 0.5368577983788596
Success in epsoide 332, used 83 iterations!
now epsilon is 0.01, the reward is -74.0 maxPosition is 0.5115580699527941
Success in epsoide 333, used 130 iterations!
now epsilon is 0.01, the reward is -121.0 maxPosition is 0.5368577983788596
Success in epsoide 334, used 138 iterations!
now epsilon is 0.01, the reward is -129.0 maxPosition is 0.5368577983788596
Success in epsoide 335, used 84 iterations!
now epsilon is 0.01, the reward is -75.0 maxPosition is 0.5169294727007353
Success in epsoide 336, used 174 iterations!
now epsilon is 0.01, the reward is -165.0 maxPosition is 0.5368577983788596
Success in epsoide 337, used 110 iterations!
now epsilon is 0.01, the reward is -101.0 maxPosition is 0.5066575500348665
Success in epsoide 338, used 112 iterations!
now epsilon is 0.01, the reward is -103.0 maxPosition is 0.5142626584307007
Success in epsoide 339, used 84 iterations!
now epsilon is 0.01, the reward is -75.0 maxPosition is 0.5056961384935496
Success in epsoide 340, used 146 iterations!
now epsilon is 0.01, the reward is -137.0 maxPosition is 0.5257126142082267
Success in epsoide 341, used 153 iterations!
now epsilon is 0.01, the reward is -144.0 maxPosition is 0.5368577983788596
Success in epsoide 342, used 118 iterations!
now epsilon is 0.01, the reward is -109.0 maxPosition is 0.5022375123349831
Success in epsoide 343, used 107 iterations!
now epsilon is 0.01, the reward is -98.0 maxPosition is 0.5368577983788596
Success in epsoide 344, used 126 iterations!
now epsilon is 0.01, the reward is -117.0 maxPosition is 0.5368577983788596
Success in epsoide 345, used 156 iterations!
now epsilon is 0.01, the reward is -147.0 maxPosition is 0.5215723838392536
Success in epsoide 346, used 115 iterations!
now epsilon is 0.01, the reward is -106.0 maxPosition is 0.5350022315419667
Success in epsoide 347, used 144 iterations!
now epsilon is 0.01, the reward is -135.0 maxPosition is 0.5129352477415754
Success in epsoide 348, used 123 iterations!
now epsilon is 0.01, the reward is -114.0 maxPosition is 0.5255893747224784
Success in epsoide 349, used 148 iterations!
now epsilon is 0.01, the reward is -139.0 maxPosition is 0.5129352477415754
Success in epsoide 350, used 113 iterations!
now epsilon is 0.01, the reward is -104.0 maxPosition is 0.5368577983788596
Success in epsoide 351, used 117 iterations!
now epsilon is 0.01, the reward is -108.0 maxPosition is 0.5219072612042372
Success in epsoide 352, used 142 iterations!
now epsilon is 0.01, the reward is -133.0 maxPosition is 0.531806603976885
Success in epsoide 353, used 87 iterations!
now epsilon is 0.01, the reward is -78.0 maxPosition is 0.5074467006083948
Success in epsoide 354, used 83 iterations!
now epsilon is 0.01, the reward is -74.0 maxPosition is 0.5094936491906414
Success in epsoide 355, used 114 iterations!
now epsilon is 0.01, the reward is -105.0 maxPosition is 0.5220428176325399
Success in epsoide 356, used 84 iterations!
now epsilon is 0.01, the reward is -75.0 maxPosition is 0.5069459992731968
Success in epsoide 357, used 108 iterations!
now epsilon is 0.01, the reward is -99.0 maxPosition is 0.5136887664481644
Success in epsoide 358, used 104 iterations!
now epsilon is 0.01, the reward is -95.0 maxPosition is 0.5068717433547105
Success in epsoide 359, used 148 iterations!
now epsilon is 0.01, the reward is -139.0 maxPosition is 0.5368577983788596
Success in epsoide 360, used 139 iterations!
now epsilon is 0.01, the reward is -130.0 maxPosition is 0.5371390279228068
Success in epsoide 361, used 169 iterations!
now epsilon is 0.01, the reward is -160.0 maxPosition is 0.5363251097334638
Success in epsoide 362, used 141 iterations!
now epsilon is 0.01, the reward is -132.0 maxPosition is 0.5387670642806129
```

```
Success in epsoide 363, used 152 iterations!
now epsilon is 0.01, the reward is -143.0 maxPosition is 0.5368577983788596
Success in epsoide 364, used 109 iterations!
now epsilon is 0.01, the reward is -100.0 maxPosition is 0.5103039789848584
Success in epsoide 365, used 153 iterations!
now epsilon is 0.01, the reward is -144.0 maxPosition is 0.5368577983788596
Success in epsoide 366, used 110 iterations!
now epsilon is 0.01, the reward is -101.0 maxPosition is 0.5331773244315756
Success in epsoide 367, used 121 iterations!
now epsilon is 0.01, the reward is -112.0 maxPosition is 0.5415510801349341
Success in epsoide 368, used 116 iterations!
now epsilon is 0.01, the reward is -107.0 maxPosition is 0.5416077270794938
Success in epsoide 369, used 107 iterations!
now epsilon is 0.01, the reward is -98.0 maxPosition is 0.5238782807602149
Success in epsoide 370, used 125 iterations!
now epsilon is 0.01, the reward is -116.0 maxPosition is 0.524843358283106
Success in epsoide 371, used 113 iterations!
now epsilon is 0.01, the reward is -104.0 maxPosition is 0.5082745260400259
Success in epsoide 372, used 109 iterations!
now epsilon is 0.01, the reward is -100.0 maxPosition is 0.5368577983788596
Success in epsoide 373, used 111 iterations!
now epsilon is 0.01, the reward is -102.0 maxPosition is 0.5334419889449963
Success in epsoide 374, used 118 iterations!
now epsilon is 0.01, the reward is -109.0 maxPosition is 0.5126415094929134
Success in epsoide 375, used 115 iterations!
now epsilon is 0.01, the reward is -106.0 maxPosition is 0.5288914748242523
Success in epsoide 376, used 170 iterations!
now epsilon is 0.01, the reward is -161.0 maxPosition is 0.5420727081567481
Success in epsoide 377, used 139 iterations!
now epsilon is 0.01, the reward is -130.0 maxPosition is 0.5368577983788596
Success in epsoide 378, used 112 iterations!
now epsilon is 0.01, the reward is -103.0 maxPosition is 0.5044215392055983
Success in epsoide 379, used 120 iterations!
now epsilon is 0.01, the reward is -111.0 maxPosition is 0.534332189475921
Success in epsoide 380, used 153 iterations!
now epsilon is 0.01, the reward is -144.0 maxPosition is 0.5171996084428573
Success in epsoide 381, used 119 iterations!
now epsilon is 0.01, the reward is -110.0 maxPosition is 0.5201790005648352
Success in epsoide 382, used 116 iterations!
now epsilon is 0.01, the reward is -107.0 maxPosition is 0.5327995356778372
Success in epsoide 383, used 118 iterations!
now epsilon is 0.01, the reward is -109.0 maxPosition is 0.5202960534529724
Success in epsoide 384, used 114 iterations!
now epsilon is 0.01, the reward is -105.0 maxPosition is 0.5263137791606775
Success in epsoide 385, used 193 iterations!
now epsilon is 0.01, the reward is -184.0 maxPosition is 0.538779943878305
Success in epsoide 386, used 135 iterations!
now epsilon is 0.01, the reward is -126.0 maxPosition is 0.5016541452092518
Success in epsoide 387, used 119 iterations!
now epsilon is 0.01, the reward is -110.0 maxPosition is 0.5121105382877043
Success in epsoide 388, used 110 iterations!
now epsilon is 0.01, the reward is -101.0 maxPosition is 0.5028131788767619
Success in epsoide 389, used 113 iterations!
now epsilon is 0.01, the reward is -104.0 maxPosition is 0.5156746093942552
Success in epsoide 390, used 112 iterations!
now epsilon is 0.01, the reward is -103.0 maxPosition is 0.5220196017047704
Success in epsoide 391, used 179 iterations!
now epsilon is 0.01, the reward is -170.0 maxPosition is 0.5368577983788596
Success in epsoide 392, used 176 iterations!
now epsilon is 0.01, the reward is -167.0 maxPosition is 0.5368577983788596
Success in epsoide 393, used 112 iterations!
now epsilon is 0.01, the reward is -103.0 maxPosition is 0.5089490934484102
Success in epsoide 394, used 111 iterations!
now epsilon is 0.01, the reward is -102.0 maxPosition is 0.5203538584174375
Success in epsoide 395, used 115 iterations!
now epsilon is 0.01, the reward is -106.0 maxPosition is 0.5253030230069918
Success in epsoide 396, used 109 iterations!
now epsilon is 0.01, the reward is -100.0 maxPosition is 0.502343098735524
Success in epsoide 397, used 123 iterations!
```

```
now epsilon is 0.01, the reward is -114.0 maxPosition is 0.506910749017576
Success in epsoide 398, used 113 iterations!
now epsilon is 0.01, the reward is -104.0 maxPosition is 0.5239921412095027
Success in epsoide 399, used 114 iterations!
now epsilon is 0.01, the reward is -105.0 maxPosition is 0.5186724733350173
```

In [ ]:

```
env = gym.make('MountainCar-v0')

#play 20 times
#load the network
model=models.load_model('trainNetworkInEPS399.h5')

completed = 0
num_episodes = 20

for i_episode in range(num_episodes):

    currentState = env.reset().reshape(1, 2)

    print("===========================================")

    rewardSum=0
    done = False
    t = 0
    while not done:
        # env.render()
        action = np.argmax(model.predict(currentState)[0])

        new_state, reward, done, info = env.step(action)

        new_state = new_state.reshape(1, 2)

        currentState=new_state

        rewardSum+=reward

        t+=1

        if t == 200 :
          print("Episode finished but couldnot reach the top of the hill")
          break

        if done:
            completed +=1
            print("Episode finished after {} timesteps reward is {}".format(t,reward
            break

print(f"Among {num_episodes} , {completed} episodes were completed and able to reach
```

```
===========================================
Episode finished after 110 timesteps reward is -110.0
===========================================
Episode finished but couldnot reach the top of the hill
===========================================
Episode finished after 122 timesteps reward is -122.0
===========================================
Episode finished but couldnot reach the top of the hill
===========================================
Episode finished after 116 timesteps reward is -116.0
===========================================
Episode finished after 118 timesteps reward is -118.0
===========================================
Episode finished after 116 timesteps reward is -116.0
===========================================
Episode finished after 110 timesteps reward is -110.0
```

```
==========================================
Episode finished after 109 timesteps reward is -109.0
==========================================
Episode finished after 116 timesteps reward is -116.0
==========================================
Episode finished after 111 timesteps reward is -111.0
==========================================
Episode finished after 112 timesteps reward is -112.0
==========================================
Episode finished after 110 timesteps reward is -110.0
==========================================
Episode finished after 115 timesteps reward is -115.0
==========================================
Episode finished after 121 timesteps reward is -121.0
==========================================
Episode finished after 116 timesteps reward is -116.0
==========================================
Episode finished after 115 timesteps reward is -115.0
==========================================
Episode finished after 115 timesteps reward is -115.0
==========================================
Episode finished after 112 timesteps reward is -112.0
==========================================
Episode finished after 112 timesteps reward is -112.0
Among 20 , 18 episodes were completed and able to reach the top of the hill
```

Roulette

In [ ]:

```
class RouletteTrain:
    def __init__(self,env):
        self.env=env
        self.gamma=0.99

        self.epsilon = 1
        self.epsilon_decay = 0.05

        self.epsilon_min=0.01


        self.learingRate=0.001

        self.replayBuffer=deque(maxlen=20000)
        self.trainNetwork=self.createNetwork()

        self.episodeNum=100

        self.iterationNum=201 #max is 200

        self.numPickFromBuffer=32

        self.targetNetwork=self.createNetwork()

        self.targetNetwork.set_weights(self.trainNetwork.get_weights())

    def createNetwork(self):
        model = models.Sequential()
        state_shape = self.env.observation_space.shape

        model.add(layers.Dense(24, activation='relu', input_shape=(1,1)))
        model.add(layers.Dense(48, activation='relu'))
        model.add(layers.Dense(self.env.action_space.n,activation='linear'))
        model.compile(loss='mse', optimizer=Adam(learning_rate=self.learingRate))
        return model
```

```
    def getBestAction(self,state):
        self.epsilon = max(self.epsilon_min, self.epsilon)

        if np.random.rand(1) < self.epsilon:
            action = np.random.randint(0, 3)
        else:
            action=np.argmax(self.trainNetwork.predict(state)[0])

        return action



    def trainFromBuffer_Boost(self):
        if len(self.replayBuffer) < self.numPickFromBuffer:
            return
        samples = random.sample(self.replayBuffer,self.numPickFromBuffer)
        npsamples = np.array(samples)
        states_temp, actions_temp, rewards_temp, newstates_temp, dones_temp = np.hsp
        states = np.concatenate((np.squeeze(states_temp[:])), axis = 0)
        rewards = rewards_temp.reshape(self.numPickFromBuffer,).astype(float)
        targets = self.trainNetwork.predict(states)
        newstates = np.concatenate(np.concatenate(newstates_temp))
        dones = np.concatenate(dones_temp).astype(bool)
        notdones = ~dones
        notdones = notdones.astype(float)
        dones = dones.astype(float)
        Q_futures = self.targetNetwork.predict(newstates).max(axis = 1)
        targets[(np.arange(self.numPickFromBuffer), actions_temp.reshape(self.numPic
        self.trainNetwork.fit(states, targets, epochs=1, verbose=0)



    def trainFromBuffer(self):
        if len(self.replayBuffer) < self.numPickFromBuffer:
            return

        samples = random.sample(self.replayBuffer,self.numPickFromBuffer)
        states = []
        newStates=[]
        for sample in samples:
            state, action, reward, new_state, done = sample
            states.append(state)
            newStates.append(new_state)

        newArray = np.array(states)
        states = newArray.reshape(self.numPickFromBuffer, 1)

        newArray2 = np.array(newStates)
        newStates = newArray2.reshape(self.numPickFromBuffer, 1)

        targets = self.trainNetwork.predict(states)
        new_state_targets=self.targetNetwork.predict(newStates)

        i=0
        for sample in samples:
            state, action, reward, new_state, done = sample
            target = targets[i]
            if done:
                target[0][action] = reward
            else:
                Q_future = max(new_state_targets[i][0])
                target[0][action] = reward + Q_future * self.gamma
            i+=1
```

```
        self.trainNetwork.fit(states, targets, epochs=1, verbose=0)


    def orginalTry(self,currentState,eps):
        rewardSum = 0

        for i in range(self.iterationNum):
            bestAction = self.getBestAction(currentState)

            nw_state, reward, done, _ = env.step(bestAction)

            new_state = np.zeros((1,1),dtype=np.float64)
            new_state[0] = nw_state

            self.replayBuffer.append([currentState, bestAction, reward, new_state, d

            #Or you can use self.trainFromBuffer_Boost(), it is a matrix wise versio
            self.trainFromBuffer()

            rewardSum += reward

            currentState = new_state

            if done:
                break

        print("Success in epsoide {}, used {} iterations!".format(eps, i))
        self.trainNetwork.save('./trainNetworkInEPS{}.h5'.format(eps))

        #Sync
        self.targetNetwork.set_weights(self.trainNetwork.get_weights())

        print("now epsilon is {}, the reward is {}".format(max(self.epsilon_min, sel
        self.epsilon -= self.epsilon_decay

    def start(self):
        for eps in range(self.episodeNum):
            a = env.reset()
            currentState=np.zeros((1,1),dtype=np.float64)
            currentState[0] = a
            self.orginalTry(currentState, eps)
```

In [ ]:

```
env = gym.make('Roulette-v0')
dqn=RouletteTrain(env=env)
dqn.start()
```

```
Success in epsoide 0, used 99 iterations!
now epsilon is 1, the reward is 13.0
Success in epsoide 1, used 99 iterations!
now epsilon is 0.95, the reward is 1.0
Success in epsoide 2, used 99 iterations!
now epsilon is 0.8999999999999999, the reward is -32.0
Success in epsoide 3, used 99 iterations!
now epsilon is 0.8499999999999999, the reward is -28.0
Success in epsoide 4, used 99 iterations!
now epsilon is 0.7999999999999998, the reward is -8.0
Success in epsoide 5, used 99 iterations!
now epsilon is 0.7499999999999998, the reward is -32.0
Success in epsoide 6, used 99 iterations!
now epsilon is 0.6999999999999997, the reward is 13.0
Success in epsoide 7, used 99 iterations!
now epsilon is 0.6499999999999997, the reward is -6.0
Success in epsoide 8, used 99 iterations!
now epsilon is 0.5999999999999996, the reward is -8.0
```

```
Success in epsoide 9, used 99 iterations!
now epsilon is 0.5499999999999996, the reward is -8.0
Success in epsoide 10, used 99 iterations!
now epsilon is 0.4999999999999996, the reward is 44.0
Success in epsoide 11, used 99 iterations!
now epsilon is 0.4499999999999996, the reward is 15.0
Success in epsoide 12, used 99 iterations!
now epsilon is 0.39999999999999963, the reward is -18.0
Success in epsoide 13, used 99 iterations!
now epsilon is 0.34999999999999964, the reward is 21.0
Success in epsoide 14, used 99 iterations!
now epsilon is 0.29999999999999966, the reward is -12.0
Success in epsoide 15, used 99 iterations!
now epsilon is 0.24999999999999967, the reward is -12.0
Success in epsoide 16, used 99 iterations!
now epsilon is 0.19999999999999968, the reward is 45.0
Success in epsoide 17, used 99 iterations!
now epsilon is 0.1499999999999997, the reward is -12.0
Success in epsoide 18, used 99 iterations!
now epsilon is 0.09999999999999969, the reward is 12.0
Success in epsoide 19, used 99 iterations!
now epsilon is 0.049999999999999684, the reward is 6.0
Success in epsoide 20, used 99 iterations!
now epsilon is 0.01, the reward is -8.0
Success in epsoide 21, used 99 iterations!
now epsilon is 0.01, the reward is 10.0
Success in epsoide 22, used 99 iterations!
now epsilon is 0.01, the reward is -6.0
Success in epsoide 23, used 99 iterations!
now epsilon is 0.01, the reward is -22.0
Success in epsoide 24, used 99 iterations!
now epsilon is 0.01, the reward is 8.0
Success in epsoide 25, used 99 iterations!
now epsilon is 0.01, the reward is 8.0
Success in epsoide 26, used 99 iterations!
now epsilon is 0.01, the reward is -12.0
Success in epsoide 27, used 99 iterations!
now epsilon is 0.01, the reward is 10.0
Success in epsoide 28, used 99 iterations!
now epsilon is 0.01, the reward is -8.0
Success in epsoide 29, used 99 iterations!
now epsilon is 0.01, the reward is -6.0
Success in epsoide 30, used 99 iterations!
now epsilon is 0.01, the reward is -4.0
Success in epsoide 31, used 99 iterations!
now epsilon is 0.01, the reward is 0.0
Success in epsoide 32, used 99 iterations!
now epsilon is 0.01, the reward is -14.0
Success in epsoide 33, used 99 iterations!
now epsilon is 0.01, the reward is -10.0
Success in epsoide 34, used 99 iterations!
now epsilon is 0.01, the reward is -10.0
Success in epsoide 35, used 99 iterations!
now epsilon is 0.01, the reward is -16.0
Success in epsoide 36, used 99 iterations!
now epsilon is 0.01, the reward is 2.0
Success in epsoide 37, used 99 iterations!
now epsilon is 0.01, the reward is -8.0
Success in epsoide 38, used 99 iterations!
now epsilon is 0.01, the reward is -12.0
Success in epsoide 39, used 99 iterations!
now epsilon is 0.01, the reward is -2.0
Success in epsoide 40, used 99 iterations!
now epsilon is 0.01, the reward is -10.0
Success in epsoide 41, used 99 iterations!
now epsilon is 0.01, the reward is -4.0
Success in epsoide 42, used 99 iterations!
now epsilon is 0.01, the reward is 8.0
Success in epsoide 43, used 99 iterations!
```

```
now epsilon is 0.01, the reward is -4.0
Success in epsoide 44, used 99 iterations!
now epsilon is 0.01, the reward is 4.0
Success in epsoide 45, used 99 iterations!
now epsilon is 0.01, the reward is 0.0
Success in epsoide 46, used 99 iterations!
now epsilon is 0.01, the reward is -10.0
Success in epsoide 47, used 99 iterations!
now epsilon is 0.01, the reward is 18.0
Success in epsoide 48, used 99 iterations!
now epsilon is 0.01, the reward is 0.0
Success in epsoide 49, used 99 iterations!
now epsilon is 0.01, the reward is -8.0
Success in epsoide 50, used 99 iterations!
now epsilon is 0.01, the reward is 10.0
Success in epsoide 51, used 99 iterations!
now epsilon is 0.01, the reward is 0.0
Success in epsoide 52, used 99 iterations!
now epsilon is 0.01, the reward is 14.0
Success in epsoide 53, used 99 iterations!
now epsilon is 0.01, the reward is 2.0
Success in epsoide 54, used 99 iterations!
now epsilon is 0.01, the reward is -2.0
Success in epsoide 55, used 99 iterations!
now epsilon is 0.01, the reward is -10.0
Success in epsoide 56, used 99 iterations!
now epsilon is 0.01, the reward is -14.0
Success in epsoide 57, used 99 iterations!
now epsilon is 0.01, the reward is -8.0
Success in epsoide 58, used 99 iterations!
now epsilon is 0.01, the reward is -4.0
Success in epsoide 59, used 99 iterations!
now epsilon is 0.01, the reward is -4.0
Success in epsoide 60, used 99 iterations!
now epsilon is 0.01, the reward is 2.0
Success in epsoide 61, used 99 iterations!
now epsilon is 0.01, the reward is 0.0
Success in epsoide 62, used 99 iterations!
now epsilon is 0.01, the reward is -2.0
Success in epsoide 63, used 99 iterations!
now epsilon is 0.01, the reward is 6.0
Success in epsoide 64, used 99 iterations!
now epsilon is 0.01, the reward is 0.0
Success in epsoide 65, used 99 iterations!
now epsilon is 0.01, the reward is -18.0
Success in epsoide 66, used 99 iterations!
now epsilon is 0.01, the reward is -14.0
Success in epsoide 67, used 99 iterations!
now epsilon is 0.01, the reward is -20.0
Success in epsoide 68, used 99 iterations!
now epsilon is 0.01, the reward is 6.0
Success in epsoide 69, used 99 iterations!
now epsilon is 0.01, the reward is 8.0
Success in epsoide 70, used 99 iterations!
now epsilon is 0.01, the reward is -6.0
Success in epsoide 71, used 99 iterations!
now epsilon is 0.01, the reward is -16.0
Success in epsoide 72, used 99 iterations!
now epsilon is 0.01, the reward is -6.0
Success in epsoide 73, used 99 iterations!
now epsilon is 0.01, the reward is -10.0
Success in epsoide 74, used 99 iterations!
now epsilon is 0.01, the reward is 16.0
Success in epsoide 75, used 99 iterations!
now epsilon is 0.01, the reward is -6.0
Success in epsoide 76, used 99 iterations!
now epsilon is 0.01, the reward is -18.0
Success in epsoide 77, used 99 iterations!
now epsilon is 0.01, the reward is 8.0
```

```
Success in epsoide 78, used 99 iterations!
now epsilon is 0.01, the reward is 12.0
Success in epsoide 79, used 99 iterations!
now epsilon is 0.01, the reward is 14.0
Success in epsoide 80, used 99 iterations!
now epsilon is 0.01, the reward is -18.0
Success in epsoide 81, used 99 iterations!
now epsilon is 0.01, the reward is 2.0
Success in epsoide 82, used 99 iterations!
now epsilon is 0.01, the reward is -12.0
Success in epsoide 83, used 99 iterations!
now epsilon is 0.01, the reward is 4.0
Success in epsoide 84, used 99 iterations!
now epsilon is 0.01, the reward is -12.0
Success in epsoide 85, used 99 iterations!
now epsilon is 0.01, the reward is 10.0
Success in epsoide 86, used 99 iterations!
now epsilon is 0.01, the reward is -6.0
Success in epsoide 87, used 99 iterations!
now epsilon is 0.01, the reward is 20.0
Success in epsoide 88, used 99 iterations!
now epsilon is 0.01, the reward is -10.0
Success in epsoide 89, used 99 iterations!
now epsilon is 0.01, the reward is -12.0
Success in epsoide 90, used 99 iterations!
now epsilon is 0.01, the reward is -4.0
Success in epsoide 91, used 99 iterations!
now epsilon is 0.01, the reward is -2.0
Success in epsoide 92, used 99 iterations!
now epsilon is 0.01, the reward is 12.0
Success in epsoide 93, used 99 iterations!
now epsilon is 0.01, the reward is -14.0
Success in epsoide 94, used 99 iterations!
now epsilon is 0.01, the reward is 0.0
Success in epsoide 95, used 99 iterations!
now epsilon is 0.01, the reward is 14.0
Success in epsoide 96, used 99 iterations!
now epsilon is 0.01, the reward is 8.0
Success in epsoide 97, used 99 iterations!
now epsilon is 0.01, the reward is 2.0
Success in epsoide 98, used 99 iterations!
now epsilon is 0.01, the reward is 8.0
Success in epsoide 99, used 99 iterations!
now epsilon is 0.01, the reward is -4.0
```

In [ ]:

```
# After training it we save those models in which we now take the latest model to us

env = gym.make('Roulette-v0')

#play 20 times
#load the network
model=models.load_model('trainNetworkInEPS99.h5')

num_episodes = 20
totalReward = 0

for i_episode in range(num_episodes):

    a = env.reset()
    currentState=np.zeros((1,1),dtype=np.float64)
    currentState[0] = a

    print("==========================================")

    rewardSum=0
    done = False
    t = 0
```

```
    while not done:
        # env.render()
        action = np.argmax(model.predict(currentState)[0])

        nw_state, reward, done, info = env.step(action)

        new_state = np.zeros((1,1),dtype=np.float64)
        new_state[0] = nw_state

        currentState=new_state

        rewardSum+=reward

        t+=1

        if done:
            totalReward += rewardSum
            print("Episode finished after {} timesteps reward is {}".format(t,reward
            break


avg_rewards = int(totalReward/num_episodes)
print("Average reward points in {} episodes is {}".format(num_episodes,avg_rewards))
```

```
==========================================
Episode finished after 100 timesteps reward is -22.0
==========================================
Episode finished after 100 timesteps reward is -4.0
==========================================
Episode finished after 100 timesteps reward is -8.0
==========================================
Episode finished after 100 timesteps reward is -4.0
==========================================
Episode finished after 100 timesteps reward is -2.0
==========================================
Episode finished after 100 timesteps reward is 0.0
==========================================
Episode finished after 100 timesteps reward is -14.0
==========================================
Episode finished after 100 timesteps reward is 0.0
==========================================
Episode finished after 100 timesteps reward is -16.0
==========================================
Episode finished after 100 timesteps reward is -8.0
==========================================
Episode finished after 100 timesteps reward is 10.0
==========================================
Episode finished after 100 timesteps reward is -10.0
==========================================
Episode finished after 100 timesteps reward is 4.0
==========================================
Episode finished after 100 timesteps reward is -2.0
==========================================
Episode finished after 100 timesteps reward is 4.0
==========================================
Episode finished after 100 timesteps reward is -6.0
==========================================
Episode finished after 100 timesteps reward is 2.0
==========================================
Episode finished after 100 timesteps reward is -12.0
==========================================
Episode finished after 100 timesteps reward is 4.0
==========================================
Episode finished after 100 timesteps reward is 4.0
Average reward points in 20 episodes is -4
```