

```
import gym
import numpy as np

# Creating the Cart-Pole game environment
env = gym.make('CartPole-v1')

# environment details
print('Number of possible actions:', env.action_space.n)
print('States:', env.observation_space)
print(env.observation_space.low)
print(env.observation_space.high)

Number of possible actions: 2
States: Box([-4.8000002e+00 -3.4028235e+38 -4.1887903e-01 -
3.4028235e+38], [4.8000002e+00 3.4028235e+38 4.1887903e-01
3.4028235e+38], (4,), float32)
[-4.8000002e+00 -3.4028235e+38 -4.1887903e-01 -3.4028235e+38]
[4.8000002e+00 3.4028235e+38 4.1887903e-01 3.4028235e+38]
```

In the Cart-Pole game, there are two possible actions, moving the cart left or right. So there are two actions we can take, encoded as 0 and 1.

```
from time import sleep

actions = [] # actions that the agent selects
rewards = [] # obtained rewards
state = env.reset()

while True:
    action = env.action_space.sample() # choose a random action
    result = env.step(action)
    state, reward, done, _ = result[:4] # Use only the first four values
    env.render()
    sleep(0.15)
    rewards.append(reward)
    actions.append(action)
    if done:
        env.close()
        break

print('Actions:', actions)
print('Rewards:', rewards)
```

Actions: [0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0,
0, 1, 0, 1, 1]

Rewards: [1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0,
1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0]

```

import tensorflow as tf
tf.compat.v1.disable_eager_execution()

def fully_connected(name, input_tensor, num_units,
                    activation=tf.nn.relu):
    """Returns a fully connected layer"""
    with tf.compat.v1.variable_scope(name,
        reuse=tf.compat.v1.AUTO_REUSE):
        # initialize weights
        w = tf.compat.v1.get_variable(f"W_{name}",
            shape=[input_tensor.get_shape()[1], num_units],
            initializer=tf.compat.v1.initializers.he_uniform(),
                                dtype=tf.float32,
            trainable=True)
        # initialize bias
        b = tf.compat.v1.get_variable(f"B_{name}", shape=[num_units],
            initializer=tf.constant_initializer(0.0),
                                dtype=tf.float32,
                                trainable=True)

        # output
        out = tf.matmul(input_tensor, w) + b
        # add activation
        if activation:
            out = activation(out, name=f"activation_{name}")
        # change name
        out = tf.compat.v1.identity(out, name=name)

    return out

class QNetwork:
    def __init__(self, learning_rate=0.01, state_size=4,
                 action_size=2, hidden_size=10,
                 name='QNetwork'):
        with tf.compat.v1.variable_scope(name):
            # state inputs to the Q-network
            self.inputs_ = tf.compat.v1.placeholder(tf.float32, [None,
                state_size], name='inputs')

            # One hot encode the actions to later choose the Q-value
            for the action
            self.actions_ = tf.compat.v1.placeholder(tf.int32, [None],
                name='actions')
            one_hot_actions = tf.one_hot(self.actions_, action_size)

            # Target Q values for training
            self.targetQs_ = tf.compat.v1.placeholder(tf.float32,
                [None], name='target')

```

```

        # ReLU hidden layers
        self.fc1 = fully_connected("h1", self.inputs_,
hidden_size)
        self.fc2 = fully_connected("h2", self.fc1, hidden_size)
        self.fc3 = fully_connected("h3", self.fc2, hidden_size)

        # Linear output layer
        self.output = fully_connected("output", self.fc3,
action_size, activation=None)

        ### Train using mean squared error and Adam gradient
descent.
        self.Q = tf.reduce_sum(tf.multiply(self.output,
one_hot_actions), axis=1)
        self.loss = tf.reduce_mean(tf.square(self.targetQs_ -
self.Q))
        self.opt =
tf.compat.v1.train.AdamOptimizer(learning_rate).minimize(self.loss)

```

Experience replay

```

from collections import deque

class Memory():
    def __init__(self, max_size=1000):
        self.buffer = deque(maxlen=max_size)

    def add(self, experience):
        self.buffer.append(experience)

    def sample(self, batch_size):
        idx = np.random.choice(np.arange(len(self.buffer)),
                                size=batch_size,
                                replace=False)
        return [self.buffer[ii] for ii in idx]

```

Hyperparameters

```

train_episodes = 700          # max number of episodes to learn from
max_steps = 500               # max steps in an episode
gamma = 1.0                   # future reward discount

# Exploration parameters
explore_start = 1.0           # exploration probability at start
explore_stop = 0.01           # minimum exploration probability
decay_rate = 0.0002           # exponential decay rate for
exploration prob

# Network parameters

```

```

hidden_size = 64                # number of units in each Q-network
hidden_layer
learning_rate = 0.0001          # Q-network learning rate

# Memory parameters
memory_size = 100000            # memory capacity
batch_size = 32                 # experience mini-batch size
pretrain_length = batch_size    # number experiences to pretrain the
memory

mainQN = QNetwork(name='main', hidden_size=hidden_size,
learning_rate=learning_rate)

WARNING:tensorflow:From C:\Users\Shubham\AppData\Local\Temp\
ipykernel_26820\2245679123.py:9: The name tf.initializers.he_uniform
is deprecated. Please use tf.compat.v1.initializers.he_uniform
instead.

```

Populate the experience memory

```

# Initialize the environment
env.reset()

memory = Memory(max_size=memory_size)

# Make a bunch of random actions and store the experiences
for ii in range(pretrain_length):

    # Make a random action
    action = env.action_space.sample()
    result = env.step(action)

    # Check if the result is a tuple and has at least four elements
    if isinstance(result, tuple) and len(result) >= 4:
        state, reward, done, info = result[:4]

        # reduce the reward if the cart is not centered
        reward = max(0, reward * (1 - abs(state[0]/2.4)))

        if done:
            # The simulation fails, so no next state
            next_state = np.zeros(state.shape)
            # Add experience to memory
            memory.add((state, action, reward, next_state))

            # Start a new episode
            env.reset()
        else:
            # Add experience to memory

```

```

        memory.add((state, action, reward, state)) # Assuming
state is the next_state
    else:
        raise ValueError("Unexpected result from env.step():
{}".format(result))

```

Training

```

import tensorflow as tf
import numpy as np
from time import sleep
from collections import deque

# Define your QNetwork and Memory classes here if not defined already

class QNetwork(tf.keras.Model):
    def __init__(self, learning_rate=0.01, state_size=4,
action_size=2, hidden_size=10, name='QNetwork'):
        super(QNetwork, self).__init__(name=name)
        # ... (your QNetwork class definition)
# Create a QNetwork instance
mainQN = QNetwork()

# Create a Checkpoint object
checkpoint = tf.train.Checkpoint(model=mainQN)

# Initialize the environment
env.reset()

# ... (your other configurations)

rewards_list = []

# Define a saver object
saver = tf.compat.v1.train.Saver()

with tf.compat.v1.Session() as sess:
    # Initialize variables
    sess.run(tf.compat.v1.global_variables_initializer())

    step = 0
    for ep in range(1, train_episodes):
        total_reward = 0
        t = 0
        while t < max_steps:
            step += 1
            env.render()
            sleep(0.0005)

            explore_p = explore_stop + (explore_start - explore_stop)

```

```

* np.exp(-decay_rate * step)
    if explore_p > np.random.rand():
        action = env.action_space.sample()
    else:
        feed = {mainQN.inputs_: state.reshape((1,
*state.shape))}
        Qs = sess.run(mainQN.output, feed_dict=feed)
        action = np.argmax(Qs)

    result = env.step(action)
    next_state, reward, done, _ = result[0], result[1],
result[2], result[3]

    reward = max(0, reward * (1 - abs(next_state[0] / 2.4)))
    total_reward += reward

    if done:
        next_state = np.zeros(state.shape)
        t = max_steps

        print('Episode: {}'.format(ep),
              'Total reward: {}'.format(total_reward),
              'Explore P: {:.4f}'.format(explore_p))
        rewards_list.append((ep, total_reward))

        memory.add((state, action, reward, next_state))

        env.reset()
        result = env.step(env.action_space.sample())
        state, reward, done, _ = result[0], result[1],
result[2], result[3]

    else:
        memory.add((state, action, reward, next_state))
        state = next_state
        t += 1

    batch = memory.sample(batch_size)
    states = np.array([each[0] for each in batch])
    actions = np.array([each[1] for each in batch])
    rewards = np.array([each[2] for each in batch])
    next_states = np.array([each[3] for each in batch])

    target_Qs = sess.run(mainQN.output,
feed_dict={mainQN.inputs_: next_states})
    episode_ends = (next_states ==
np.zeros(states[0].shape)).all(axis=1)
    target_Qs[episode_ends] = (0, 0)
    targets = rewards + gamma * np.max(target_Qs, axis=1)

```

```

        loss, _ = sess.run([mainQN.loss, mainQN.opt],
                            feed_dict={mainQN.inputs_: states,
                                        mainQN.targetQs_: targets,
                                        mainQN.actions_: actions})

    if len(rewards_list) > 10:
        stop_training = False
        for reward in rewards_list[-10:]:
            if reward[1] < 470:
                break
        else:
            stop_training = True
        if stop_training:
            break

    # Save the model at each episode
    saver.save(sess,
"C:/Users/Shubham/Downloads/.ipynb_checkpoints/CartPole.ckpt")

```

```

Episode: 1 Total reward: 15.74248934459562 Explore P: 0.9968
Episode: 2 Total reward: 14.766307852478349 Explore P: 0.9939
Episode: 3 Total reward: 15.756548851728438 Explore P: 0.9907
Episode: 4 Total reward: 18.78910342871435 Explore P: 0.9870
Episode: 5 Total reward: 59.262726108863724 Explore P: 0.9744
Episode: 6 Total reward: 48.87488476838917 Explore P: 0.9646
Episode: 7 Total reward: 13.81380879714076 Explore P: 0.9619
Episode: 8 Total reward: 41.61779662159581 Explore P: 0.9538
Episode: 9 Total reward: 14.750033543678002 Explore P: 0.9510
Episode: 10 Total reward: 17.481490585021675 Explore P: 0.9476
Episode: 11 Total reward: 36.08977787029774 Explore P: 0.9407
Episode: 12 Total reward: 15.726056907044647 Explore P: 0.9377
Episode: 13 Total reward: 15.76177281991113 Explore P: 0.9347
Episode: 14 Total reward: 11.73770309132912 Explore P: 0.9325
Episode: 15 Total reward: 9.779315360550148 Explore P: 0.9307
Episode: 16 Total reward: 22.07815465936437 Explore P: 0.9264
Episode: 17 Total reward: 24.350596914979786 Explore P: 0.9219
Episode: 18 Total reward: 24.821968359271224 Explore P: 0.9173
Episode: 19 Total reward: 12.679311816580592 Explore P: 0.9150
Episode: 20 Total reward: 12.503851315627495 Explore P: 0.9126
Episode: 21 Total reward: 29.187542079792667 Explore P: 0.9072
Episode: 22 Total reward: 23.136875641027775 Explore P: 0.9029
Episode: 23 Total reward: 33.80766221376446 Explore P: 0.8969
Episode: 24 Total reward: 12.501156648620963 Explore P: 0.8946
Episode: 25 Total reward: 34.36661322688452 Explore P: 0.8884
Episode: 26 Total reward: 52.441809941859304 Explore P: 0.8788
Episode: 27 Total reward: 16.78103840975867 Explore P: 0.8758
Episode: 28 Total reward: 20.74863065326645 Explore P: 0.8722
Episode: 29 Total reward: 46.41804367646304 Explore P: 0.8641
Episode: 30 Total reward: 9.74396256186689 Explore P: 0.8624
Episode: 31 Total reward: 11.547598830734689 Explore P: 0.8604

```

Episode: 32 Total reward: 16.660862427602602 Explore P: 0.8575
Episode: 33 Total reward: 10.645961709320547 Explore P: 0.8556
Episode: 34 Total reward: 12.567488245666027 Explore P: 0.8535
Episode: 35 Total reward: 35.20275368373646 Explore P: 0.8474
Episode: 36 Total reward: 22.286095034331083 Explore P: 0.8436
Episode: 37 Total reward: 14.804730063478928 Explore P: 0.8411
Episode: 38 Total reward: 16.60949871128832 Explore P: 0.8382
Episode: 39 Total reward: 11.657655088910056 Explore P: 0.8363
Episode: 40 Total reward: 10.69632624144045 Explore P: 0.8344
Episode: 41 Total reward: 21.517185141545877 Explore P: 0.8308
Episode: 42 Total reward: 17.515184889314696 Explore P: 0.8279
Episode: 43 Total reward: 34.7356048921356 Explore P: 0.8220
Episode: 44 Total reward: 19.794982699677348 Explore P: 0.8188
Episode: 45 Total reward: 10.82217163603006 Explore P: 0.8170
Episode: 46 Total reward: 12.660560519744953 Explore P: 0.8149
Episode: 47 Total reward: 10.651487574446946 Explore P: 0.8131
Episode: 48 Total reward: 8.747023655218072 Explore P: 0.8117
Episode: 49 Total reward: 19.331076553983923 Explore P: 0.8085
Episode: 50 Total reward: 8.607775261004766 Explore P: 0.8070
Episode: 51 Total reward: 18.665957324662788 Explore P: 0.8040
Episode: 52 Total reward: 7.842445830775736 Explore P: 0.8027
Episode: 53 Total reward: 9.564875642148156 Explore P: 0.8012
Episode: 54 Total reward: 13.785056931626363 Explore P: 0.7990
Episode: 55 Total reward: 9.640185394479582 Explore P: 0.7974
Episode: 56 Total reward: 11.387332747379938 Explore P: 0.7955
Episode: 57 Total reward: 31.57418446118633 Explore P: 0.7905
Episode: 58 Total reward: 31.528913395964384 Explore P: 0.7855
Episode: 59 Total reward: 7.879911931930111 Explore P: 0.7843
Episode: 60 Total reward: 33.72688488477919 Explore P: 0.7790
Episode: 61 Total reward: 9.808960986362461 Explore P: 0.7775
Episode: 62 Total reward: 19.333945275672406 Explore P: 0.7744
Episode: 63 Total reward: 16.271887661889195 Explore P: 0.7718
Episode: 64 Total reward: 21.110999793047082 Explore P: 0.7685
Episode: 65 Total reward: 20.44370921390752 Explore P: 0.7653
Episode: 66 Total reward: 16.518157962399226 Explore P: 0.7627
Episode: 67 Total reward: 14.596938680236537 Explore P: 0.7605
Episode: 68 Total reward: 13.498114538534235 Explore P: 0.7584
Episode: 69 Total reward: 14.351544855162503 Explore P: 0.7561
Episode: 70 Total reward: 8.880475062661086 Explore P: 0.7548
Episode: 71 Total reward: 11.80546111454411 Explore P: 0.7530
Episode: 72 Total reward: 15.41765629996856 Explore P: 0.7506
Episode: 73 Total reward: 13.781264090973613 Explore P: 0.7486
Episode: 74 Total reward: 13.835952662251659 Explore P: 0.7465
Episode: 75 Total reward: 13.827320269929867 Explore P: 0.7444
Episode: 76 Total reward: 22.49434474886706 Explore P: 0.7411
Episode: 77 Total reward: 48.06769405118636 Explore P: 0.7335
Episode: 78 Total reward: 10.825557625232857 Explore P: 0.7319
Episode: 79 Total reward: 38.22684036015805 Explore P: 0.7263
Episode: 80 Total reward: 8.80211048872055 Explore P: 0.7250

Episode: 81 Total reward: 12.616955509409308 Explore P: 0.7232
Episode: 82 Total reward: 9.800580002483912 Explore P: 0.7217
Episode: 83 Total reward: 12.728785800980404 Explore P: 0.7199
Episode: 84 Total reward: 8.601271188197035 Explore P: 0.7186
Episode: 85 Total reward: 23.173860689625144 Explore P: 0.7152
Episode: 86 Total reward: 39.930420000809434 Explore P: 0.7095
Episode: 87 Total reward: 11.579102591766667 Explore P: 0.7078
Episode: 88 Total reward: 13.805449608174968 Explore P: 0.7058
Episode: 89 Total reward: 47.92653523353026 Explore P: 0.6988
Episode: 90 Total reward: 10.776453915983438 Explore P: 0.6973
Episode: 91 Total reward: 11.444367460285621 Explore P: 0.6956
Episode: 92 Total reward: 9.827877614491928 Explore P: 0.6942
Episode: 93 Total reward: 9.51491146348417 Explore P: 0.6929
Episode: 94 Total reward: 20.359304822961956 Explore P: 0.6900
Episode: 95 Total reward: 10.819758902691925 Explore P: 0.6885
Episode: 96 Total reward: 22.83580299715201 Explore P: 0.6853
Episode: 97 Total reward: 13.68871897047696 Explore P: 0.6834
Episode: 98 Total reward: 39.91104839919474 Explore P: 0.6777
Episode: 99 Total reward: 16.73291151332281 Explore P: 0.6755
Episode: 100 Total reward: 13.60283314071906 Explore P: 0.6736
Episode: 101 Total reward: 18.36751651732872 Explore P: 0.6711
Episode: 102 Total reward: 9.835114789457293 Explore P: 0.6698
Episode: 103 Total reward: 10.577803839929402 Explore P: 0.6683
Episode: 104 Total reward: 11.695414075492105 Explore P: 0.6668
Episode: 105 Total reward: 18.73792128129086 Explore P: 0.6643
Episode: 106 Total reward: 10.432086362503469 Explore P: 0.6628
Episode: 107 Total reward: 8.806153924922304 Explore P: 0.6616
Episode: 108 Total reward: 20.629580935967777 Explore P: 0.6589
Episode: 109 Total reward: 17.510244226548824 Explore P: 0.6566
Episode: 110 Total reward: 41.47516990675407 Explore P: 0.6510
Episode: 111 Total reward: 10.821655355238667 Explore P: 0.6496
Episode: 112 Total reward: 17.57838781090686 Explore P: 0.6473
Episode: 113 Total reward: 20.574696128305128 Explore P: 0.6447
Episode: 114 Total reward: 22.453014579291146 Explore P: 0.6418
Episode: 115 Total reward: 7.872003170118357 Explore P: 0.6407
Episode: 116 Total reward: 14.477886989091834 Explore P: 0.6389
Episode: 117 Total reward: 17.694762002667023 Explore P: 0.6366
Episode: 118 Total reward: 13.739730558455143 Explore P: 0.6348
Episode: 119 Total reward: 14.57470174335564 Explore P: 0.6330
Episode: 120 Total reward: 7.837414102318386 Explore P: 0.6320
Episode: 121 Total reward: 16.54584237188101 Explore P: 0.6299
Episode: 122 Total reward: 18.80869620436958 Explore P: 0.6275
Episode: 123 Total reward: 15.727119093828758 Explore P: 0.6255
Episode: 124 Total reward: 13.652865871806473 Explore P: 0.6238
Episode: 125 Total reward: 13.570198991180709 Explore P: 0.6221
Episode: 126 Total reward: 16.774603982562745 Explore P: 0.6200
Episode: 127 Total reward: 11.813896763395556 Explore P: 0.6186
Episode: 128 Total reward: 15.611602487566413 Explore P: 0.6166
Episode: 129 Total reward: 30.641944430152456 Explore P: 0.6128

Episode: 130	Total reward: 15.33878147508949	Explore P: 0.6108
Episode: 131	Total reward: 11.88100624451666	Explore P: 0.6094
Episode: 132	Total reward: 16.75239147849304	Explore P: 0.6074
Episode: 133	Total reward: 10.762817596589835	Explore P: 0.6060
Episode: 134	Total reward: 12.693037557881325	Explore P: 0.6045
Episode: 135	Total reward: 17.758566368206324	Explore P: 0.6024
Episode: 136	Total reward: 13.68226944678463	Explore P: 0.6007
Episode: 137	Total reward: 31.321111090818395	Explore P: 0.5969
Episode: 138	Total reward: 11.759439687089376	Explore P: 0.5955
Episode: 139	Total reward: 11.5946470539396	Explore P: 0.5941
Episode: 140	Total reward: 20.520360694732513	Explore P: 0.5917
Episode: 141	Total reward: 15.423565724864604	Explore P: 0.5898
Episode: 142	Total reward: 15.475293684129912	Explore P: 0.5880
Episode: 143	Total reward: 13.43929313371579	Explore P: 0.5863
Episode: 144	Total reward: 29.77966420259327	Explore P: 0.5828
Episode: 145	Total reward: 12.60176129359752	Explore P: 0.5813
Episode: 146	Total reward: 21.41070966771804	Explore P: 0.5788
Episode: 147	Total reward: 14.750641932575187	Explore P: 0.5771
Episode: 148	Total reward: 10.732448907161599	Explore P: 0.5758
Episode: 149	Total reward: 36.54566769771433	Explore P: 0.5716
Episode: 150	Total reward: 39.501770533660114	Explore P: 0.5669
Episode: 151	Total reward: 11.59973823900024	Explore P: 0.5655
Episode: 152	Total reward: 14.779755683072533	Explore P: 0.5639
Episode: 153	Total reward: 24.707661621202234	Explore P: 0.5611
Episode: 154	Total reward: 7.864967252360657	Explore P: 0.5602
Episode: 155	Total reward: 10.722263036237564	Explore P: 0.5590
Episode: 156	Total reward: 13.495912088314068	Explore P: 0.5575
Episode: 157	Total reward: 12.792478151503019	Explore P: 0.5560
Episode: 158	Total reward: 9.721017054010494	Explore P: 0.5550
Episode: 159	Total reward: 14.692586675810158	Explore P: 0.5533
Episode: 160	Total reward: 8.594252297344307	Explore P: 0.5523
Episode: 161	Total reward: 12.812926648107048	Explore P: 0.5509
Episode: 162	Total reward: 17.54789204057306	Explore P: 0.5490
Episode: 163	Total reward: 10.673599960903328	Explore P: 0.5478
Episode: 164	Total reward: 12.677282903964322	Explore P: 0.5464
Episode: 165	Total reward: 10.715720193926245	Explore P: 0.5452
Episode: 166	Total reward: 12.652173027182775	Explore P: 0.5438
Episode: 167	Total reward: 13.384336310128369	Explore P: 0.5424
Episode: 168	Total reward: 13.473730375369392	Explore P: 0.5409
Episode: 169	Total reward: 9.693862001101177	Explore P: 0.5398
Episode: 170	Total reward: 10.836151503162304	Explore P: 0.5386
Episode: 171	Total reward: 8.814855535320628	Explore P: 0.5377
Episode: 172	Total reward: 9.732729729730638	Explore P: 0.5366
Episode: 173	Total reward: 10.471394925688703	Explore P: 0.5355
Episode: 174	Total reward: 12.547788263298571	Explore P: 0.5341
Episode: 175	Total reward: 23.77902976880917	Explore P: 0.5316
Episode: 176	Total reward: 20.620551612849038	Explore P: 0.5294
Episode: 177	Total reward: 20.578258646458075	Explore P: 0.5272
Episode: 178	Total reward: 8.675441418463985	Explore P: 0.5263

Episode: 179 Total reward: 13.725856558458569 Explore P: 0.5249
Episode: 180 Total reward: 14.47907066438347 Explore P: 0.5233
Episode: 181 Total reward: 14.605118463514374 Explore P: 0.5218
Episode: 182 Total reward: 11.586941120990861 Explore P: 0.5206
Episode: 183 Total reward: 8.584523322992029 Explore P: 0.5196
Episode: 184 Total reward: 9.73287978277464 Explore P: 0.5186
Episode: 185 Total reward: 11.764812546491157 Explore P: 0.5174
Episode: 186 Total reward: 9.839018178754484 Explore P: 0.5164
Episode: 187 Total reward: 15.445838336056719 Explore P: 0.5148
Episode: 188 Total reward: 9.764109200138288 Explore P: 0.5138
Episode: 189 Total reward: 13.686053606565112 Explore P: 0.5124
Episode: 190 Total reward: 16.62265329052267 Explore P: 0.5106
Episode: 191 Total reward: 10.536730931916585 Explore P: 0.5095
Episode: 192 Total reward: 27.739552190878996 Explore P: 0.5067
Episode: 193 Total reward: 19.519111681729555 Explore P: 0.5047
Episode: 194 Total reward: 16.714966406318126 Explore P: 0.5030
Episode: 195 Total reward: 14.57384417221571 Explore P: 0.5015
Episode: 196 Total reward: 21.57431580922275 Explore P: 0.4994
Episode: 197 Total reward: 28.61794866372657 Explore P: 0.4965
Episode: 198 Total reward: 25.311783754140688 Explore P: 0.4940
Episode: 199 Total reward: 19.103361821422975 Explore P: 0.4921
Episode: 200 Total reward: 17.955024419202044 Explore P: 0.4903
Episode: 201 Total reward: 32.17173803934323 Explore P: 0.4872
Episode: 202 Total reward: 43.40562706207857 Explore P: 0.4830
Episode: 203 Total reward: 12.437729072912285 Explore P: 0.4818
Episode: 204 Total reward: 44.65664848281692 Explore P: 0.4775
Episode: 205 Total reward: 26.64348495249094 Explore P: 0.4749
Episode: 206 Total reward: 31.832216153693786 Explore P: 0.4720
Episode: 207 Total reward: 67.68892418853646 Explore P: 0.4656
Episode: 208 Total reward: 58.437383070656026 Explore P: 0.4602
Episode: 209 Total reward: 98.12052074442909 Explore P: 0.4512
Episode: 210 Total reward: 27.28363849098484 Explore P: 0.4487
Episode: 211 Total reward: 48.1649107951671 Explore P: 0.4444
Episode: 212 Total reward: 52.357745222805534 Explore P: 0.4398
Episode: 213 Total reward: 123.55134477138442 Explore P: 0.4286
Episode: 214 Total reward: 50.06804198461275 Explore P: 0.4243
Episode: 215 Total reward: 53.395670747665775 Explore P: 0.4198
Episode: 216 Total reward: 40.830627191006585 Explore P: 0.4164
Episode: 217 Total reward: 108.21157088705033 Explore P: 0.4062
Episode: 218 Total reward: 84.55649030776115 Explore P: 0.3990
Episode: 219 Total reward: 44.48792724729477 Explore P: 0.3955
Episode: 220 Total reward: 52.73353564708182 Explore P: 0.3913
Episode: 221 Total reward: 36.54398146646176 Explore P: 0.3885
Episode: 222 Total reward: 26.532778632904712 Explore P: 0.3864
Episode: 223 Total reward: 45.12624568047854 Explore P: 0.3830
Episode: 224 Total reward: 40.67928024275049 Explore P: 0.3799
Episode: 225 Total reward: 81.79827419846154 Explore P: 0.3738
Episode: 226 Total reward: 197.41972781515983 Explore P: 0.3545
Episode: 227 Total reward: 21.784563128293183 Explore P: 0.3530

Episode: 228	Total reward: 94.45069712460828	Explore P: 0.3461
Episode: 229	Total reward: 50.01531534896638	Explore P: 0.3427
Episode: 230	Total reward: 81.59095191908398	Explore P: 0.3367
Episode: 231	Total reward: 47.42487673647702	Explore P: 0.3334
Episode: 232	Total reward: 86.56313936991117	Explore P: 0.3278
Episode: 233	Total reward: 39.602471384840705	Explore P: 0.3251
Episode: 234	Total reward: 54.24738636390732	Explore P: 0.3217
Episode: 235	Total reward: 38.04390071673938	Explore P: 0.3192
Episode: 236	Total reward: 10.711597509061296	Explore P: 0.3186
Episode: 237	Total reward: 78.1844417050015	Explore P: 0.3137
Episode: 238	Total reward: 57.08447103229504	Explore P: 0.3102
Episode: 239	Total reward: 36.512728099653025	Explore P: 0.3079
Episode: 240	Total reward: 89.53016576368128	Explore P: 0.3025
Episode: 241	Total reward: 59.71541944293617	Explore P: 0.2990
Episode: 242	Total reward: 51.27108365363285	Explore P: 0.2960
Episode: 243	Total reward: 57.223704094353394	Explore P: 0.2927
Episode: 244	Total reward: 48.868647382914794	Explore P: 0.2899
Episode: 245	Total reward: 68.54337260328855	Explore P: 0.2859
Episode: 246	Total reward: 58.924500046220295	Explore P: 0.2823
Episode: 247	Total reward: 50.43098555730345	Explore P: 0.2795
Episode: 248	Total reward: 52.94707323643768	Explore P: 0.2766
Episode: 249	Total reward: 89.08713511105454	Explore P: 0.2718
Episode: 250	Total reward: 77.69691835584672	Explore P: 0.2676
Episode: 251	Total reward: 79.49517999255721	Explore P: 0.2634
Episode: 252	Total reward: 76.54288411932748	Explore P: 0.2595
Episode: 253	Total reward: 115.82730467343445	Explore P: 0.2534
Episode: 254	Total reward: 51.265149903735924	Explore P: 0.2508
Episode: 255	Total reward: 89.47298693269356	Explore P: 0.2460
Episode: 256	Total reward: 247.52074438022106	Explore P: 0.2325
Episode: 257	Total reward: 76.32310488820079	Explore P: 0.2286
Episode: 258	Total reward: 78.3234802085814	Explore P: 0.2244
Episode: 259	Total reward: 87.67852268344718	Explore P: 0.2200
Episode: 260	Total reward: 81.11737415847894	Explore P: 0.2161
Episode: 261	Total reward: 98.72631301532968	Explore P: 0.2115
Episode: 262	Total reward: 89.77323618745628	Explore P: 0.2073
Episode: 263	Total reward: 102.98853031371253	Explore P: 0.2019
Episode: 264	Total reward: 114.40362737980706	Explore P: 0.1970
Episode: 265	Total reward: 89.96925878093073	Explore P: 0.1928
Episode: 266	Total reward: 95.86558240009857	Explore P: 0.1887
Episode: 267	Total reward: 85.43076943841882	Explore P: 0.1851
Episode: 268	Total reward: 107.45526201091712	Explore P: 0.1807
Episode: 269	Total reward: 114.2524253969556	Explore P: 0.1765
Episode: 270	Total reward: 125.18642879351684	Explore P: 0.1714
Episode: 271	Total reward: 86.69732637699535	Explore P: 0.1676
Episode: 272	Total reward: 109.31367113720627	Explore P: 0.1636
Episode: 273	Total reward: 154.53359738923967	Explore P: 0.1580
Episode: 274	Total reward: 91.19676502315272	Explore P: 0.1545
Episode: 275	Total reward: 82.06640638286865	Explore P: 0.1514
Episode: 276	Total reward: 71.72292170362681	Explore P: 0.1492

Episode: 277 Total reward: 131.3053930956133 Explore P: 0.1444
Episode: 278 Total reward: 155.4942754599883 Explore P: 0.1394
Episode: 279 Total reward: 172.40074660706648 Explore P: 0.1342
Episode: 280 Total reward: 124.78902242359864 Explore P: 0.1304
Episode: 281 Total reward: 169.0667971350776 Explore P: 0.1255
Episode: 282 Total reward: 109.38177951301142 Explore P: 0.1223
Episode: 283 Total reward: 136.34214151355744 Explore P: 0.1186
Episode: 284 Total reward: 150.14744301517925 Explore P: 0.1148
Episode: 285 Total reward: 89.18380105794252 Explore P: 0.1123
Episode: 286 Total reward: 83.96600752028942 Explore P: 0.1100
Episode: 287 Total reward: 85.65725837461645 Explore P: 0.1076
Episode: 288 Total reward: 146.38423582575828 Explore P: 0.1043
Episode: 289 Total reward: 246.50928911735545 Explore P: 0.0989
Episode: 290 Total reward: 131.6636094707816 Explore P: 0.0960
Episode: 291 Total reward: 143.01780229744813 Explore P: 0.0929
Episode: 292 Total reward: 170.56345113027425 Explore P: 0.0896
Episode: 293 Total reward: 169.78453058268246 Explore P: 0.0864
Episode: 294 Total reward: 115.26766229070194 Explore P: 0.0840
Episode: 295 Total reward: 117.42891258156551 Explore P: 0.0821
Episode: 296 Total reward: 216.0841860619718 Explore P: 0.0786
Episode: 297 Total reward: 89.61835147657743 Explore P: 0.0770
Episode: 298 Total reward: 155.01220677670125 Explore P: 0.0746
Episode: 299 Total reward: 97.39356418109188 Explore P: 0.0729
Episode: 300 Total reward: 162.83267158709347 Explore P: 0.0705
Episode: 301 Total reward: 102.42596167877005 Explore P: 0.0689
Episode: 302 Total reward: 115.31236416061681 Explore P: 0.0672
Episode: 303 Total reward: 112.93727669293486 Explore P: 0.0656
Episode: 304 Total reward: 104.94561991403923 Explore P: 0.0642
Episode: 305 Total reward: 103.60497012976944 Explore P: 0.0628
Episode: 306 Total reward: 159.90293094333418 Explore P: 0.0609
Episode: 307 Total reward: 107.75000382299669 Explore P: 0.0595
Episode: 308 Total reward: 129.75473795354327 Explore P: 0.0580
Episode: 309 Total reward: 166.53949720057904 Explore P: 0.0561
Episode: 310 Total reward: 112.79130222113358 Explore P: 0.0548
Episode: 311 Total reward: 152.43955378903772 Explore P: 0.0532
Episode: 312 Total reward: 104.78173730778991 Explore P: 0.0521
Episode: 313 Total reward: 133.17013890787103 Explore P: 0.0508
Episode: 314 Total reward: 261.29554750928946 Explore P: 0.0485
Episode: 315 Total reward: 125.71887023899279 Explore P: 0.0473
Episode: 316 Total reward: 183.20351935568868 Explore P: 0.0457
Episode: 317 Total reward: 120.76383799291754 Explore P: 0.0447
Episode: 318 Total reward: 108.71765133459121 Explore P: 0.0437
Episode: 319 Total reward: 167.01098660336476 Explore P: 0.0425
Episode: 320 Total reward: 165.46615335892173 Explore P: 0.0412
Episode: 321 Total reward: 133.18759350045005 Explore P: 0.0403
Episode: 322 Total reward: 108.37457918049763 Explore P: 0.0394
Episode: 323 Total reward: 171.02203549710987 Explore P: 0.0383
Episode: 324 Total reward: 421.77066551055776 Explore P: 0.0358
Episode: 325 Total reward: 174.7733159900332 Explore P: 0.0348

Episode: 326 Total reward: 219.60915312706487 Explore P: 0.0335
Episode: 327 Total reward: 299.3051744554735 Explore P: 0.0320
Episode: 328 Total reward: 362.5013139434598 Explore P: 0.0304
Episode: 329 Total reward: 170.40149652259424 Explore P: 0.0296
Episode: 330 Total reward: 321.705716373045 Explore P: 0.0282
Episode: 331 Total reward: 186.80696804107478 Explore P: 0.0275
Episode: 332 Total reward: 176.50217441609917 Explore P: 0.0267
Episode: 333 Total reward: 133.46693032452222 Explore P: 0.0261
Episode: 334 Total reward: 395.85867431887726 Explore P: 0.0248
Episode: 335 Total reward: 213.6801235201417 Explore P: 0.0240
Episode: 336 Total reward: 368.7040521710492 Explore P: 0.0229
Episode: 337 Total reward: 200.2941520488821 Explore P: 0.0223
Episode: 338 Total reward: 168.74085265229712 Explore P: 0.0218
Episode: 339 Total reward: 139.78559427273777 Explore P: 0.0214
Episode: 340 Total reward: 223.11579543848845 Explore P: 0.0208
Episode: 341 Total reward: 169.9510605308612 Explore P: 0.0204
Episode: 342 Total reward: 428.2455392719326 Explore P: 0.0195
Episode: 343 Total reward: 210.11434740216166 Explore P: 0.0191
Episode: 344 Total reward: 180.2027188291663 Explore P: 0.0186
Episode: 345 Total reward: 130.38476166905215 Explore P: 0.0183
Episode: 347 Total reward: 9.10470525920391 Explore P: 0.0175
Episode: 349 Total reward: 0 Explore P: 0.0168
Episode: 350 Total reward: 350.8474139757706 Explore P: 0.0163
Episode: 351 Total reward: 141.38973735428104 Explore P: 0.0160
Episode: 352 Total reward: 182.32739470107467 Explore P: 0.0158
Episode: 353 Total reward: 122.7971961523096 Explore P: 0.0156
Episode: 354 Total reward: 306.0337290912868 Explore P: 0.0152
Episode: 355 Total reward: 232.24343546764908 Explore P: 0.0149
Episode: 356 Total reward: 410.72375253496807 Explore P: 0.0145
Episode: 357 Total reward: 229.0272918680374 Explore P: 0.0143
Episode: 359 Total reward: 12.724231983224552 Explore P: 0.0138
Episode: 360 Total reward: 355.1126532231957 Explore P: 0.0136
Episode: 361 Total reward: 386.6893276137476 Explore P: 0.0132
Episode: 362 Total reward: 240.42004890966612 Explore P: 0.0131
Episode: 363 Total reward: 374.0598455970131 Explore P: 0.0128
Episode: 364 Total reward: 163.51809457810305 Explore P: 0.0127
Episode: 365 Total reward: 295.5581092995903 Explore P: 0.0125
Episode: 366 Total reward: 197.73108347324902 Explore P: 0.0124
Episode: 367 Total reward: 219.18728693667796 Explore P: 0.0123
Episode: 368 Total reward: 359.5527273224044 Explore P: 0.0121
Episode: 369 Total reward: 321.6081459052512 Explore P: 0.0119
Episode: 370 Total reward: 332.7785035018243 Explore P: 0.0118
Episode: 373 Total reward: 65.12708154817423 Explore P: 0.0114
Episode: 375 Total reward: 0.2766854166984556 Explore P: 0.0113
Episode: 377 Total reward: 174.28508678326995 Explore P: 0.0111
Episode: 379 Total reward: 22.274737298488617 Explore P: 0.0110
Episode: 380 Total reward: 357.4937921757811 Explore P: 0.0109
Episode: 382 Total reward: 13.662837247053782 Explore P: 0.0108
Episode: 385 Total reward: 211.68667635632076 Explore P: 0.0106

Episode: 386 Total reward: 199.35488648072365 Explore P: 0.0106
Episode: 388 Total reward: 120.33955558761959 Explore P: 0.0105
Episode: 390 Total reward: 102.1321556046605 Explore P: 0.0105
Episode: 391 Total reward: 228.39556993870048 Explore P: 0.0104
Episode: 392 Total reward: 343.76444975449715 Explore P: 0.0104
Episode: 395 Total reward: 28.709742218255986 Explore P: 0.0103
Episode: 396 Total reward: 330.512856578541 Explore P: 0.0103
Episode: 399 Total reward: 36.7734408477942 Explore P: 0.0102
Episode: 401 Total reward: 365.72503341734387 Explore P: 0.0102
Episode: 402 Total reward: 341.63947401319945 Explore P: 0.0102
Episode: 415 Total reward: 81.5148564676444 Explore P: 0.0101
Episode: 417 Total reward: 126.79487208525337 Explore P: 0.0100
Episode: 419 Total reward: 2.1359680891036965 Explore P: 0.0100
Episode: 421 Total reward: 7.470404128233585 Explore P: 0.0100
Episode: 423 Total reward: 4.4231957097848245 Explore P: 0.0100
Episode: 426 Total reward: 14.82309508323669 Explore P: 0.0100
Episode: 428 Total reward: 188.8735329210757 Explore P: 0.0100

Episode: 432 Total reward: 1.0740478634834278 Explore P: 0.0100
Episode: 436 Total reward: 1.4285320639610275 Explore P: 0.0100
Episode: 440 Total reward: 101.58049832781153 Explore P: 0.0100
Episode: 445 Total reward: 99.65410375595094 Explore P: 0.0100
Episode: 450 Total reward: 36.00368752082191 Explore P: 0.0100
Episode: 461 Total reward: 65.04672799507776 Explore P: 0.0100
Episode: 463 Total reward: 60.794168204069145 Explore P: 0.0100
Episode: 465 Total reward: 47.83911314606666 Explore P: 0.0100
Episode: 467 Total reward: 49.629121422767646 Explore P: 0.0100
Episode: 469 Total reward: 77.94181238611539 Explore P: 0.0100
Episode: 471 Total reward: 211.97123137613156 Explore P: 0.0100
Episode: 473 Total reward: 107.96298832197984 Explore P: 0.0100
Episode: 475 Total reward: 94.9835294485093 Explore P: 0.0100
Episode: 477 Total reward: 18.81649274627367 Explore P: 0.0100
Episode: 480 Total reward: 9.579697181781132 Explore P: 0.0100
Episode: 482 Total reward: 100.49963255226614 Explore P: 0.0100
Episode: 486 Total reward: 14.535475303729372 Explore P: 0.0100
Episode: 488 Total reward: 40.258634239435196 Explore P: 0.0100
Episode: 490 Total reward: 107.09245967864983 Explore P: 0.0100
Episode: 492 Total reward: 3.7936911880969975 Explore P: 0.0100
Episode: 494 Total reward: 155.95965347687414 Explore P: 0.0100
Episode: 497 Total reward: 174.3150052502752 Explore P: 0.0100
Episode: 499 Total reward: 149.2547595600288 Explore P: 0.0100
Episode: 501 Total reward: 206.938141440352 Explore P: 0.0100
Episode: 503 Total reward: 30.142035235961263 Explore P: 0.0100
Episode: 506 Total reward: 28.060231506824497 Explore P: 0.0100
Episode: 508 Total reward: 222.6733605153859 Explore P: 0.0100
Episode: 510 Total reward: 97.82041345536707 Explore P: 0.0100
Episode: 512 Total reward: 8.22836647431056 Explore P: 0.0100
Episode: 514 Total reward: 49.65296287337941 Explore P: 0.0100
Episode: 515 Total reward: 288.19363650415676 Explore P: 0.0100
Episode: 517 Total reward: 68.44373426834746 Explore P: 0.0100

Episode: 519 Total reward: 43.04393635193504 Explore P: 0.0100
Episode: 521 Total reward: 147.76708914339528 Explore P: 0.0100
Episode: 523 Total reward: 162.75050257643056 Explore P: 0.0100
Episode: 525 Total reward: 90.52762329081686 Explore P: 0.0100
Episode: 527 Total reward: 15.208421806494396 Explore P: 0.0100
Episode: 529 Total reward: 35.00326396028199 Explore P: 0.0100
Episode: 531 Total reward: 14.67154050866763 Explore P: 0.0100
Episode: 532 Total reward: 248.94122411258294 Explore P: 0.0100
Episode: 533 Total reward: 319.16815281942183 Explore P: 0.0100
Episode: 534 Total reward: 299.2153350872302 Explore P: 0.0100
Episode: 536 Total reward: 66.56434195240337 Explore P: 0.0100
Episode: 538 Total reward: 18.093569864829377 Explore P: 0.0100
Episode: 539 Total reward: 279.59730657771036 Explore P: 0.0100
Episode: 541 Total reward: 81.83415559927623 Explore P: 0.0100
Episode: 543 Total reward: 176.73901925235984 Explore P: 0.0100
Episode: 546 Total reward: 161.78767665723961 Explore P: 0.0100
Episode: 549 Total reward: 5.413826743761697 Explore P: 0.0100
Episode: 551 Total reward: 256.0556383207438 Explore P: 0.0100
Episode: 553 Total reward: 238.75389427319183 Explore P: 0.0100
Episode: 555 Total reward: 0.6174279650052381 Explore P: 0.0100
Episode: 557 Total reward: 4.7548941870530435 Explore P: 0.0100
Episode: 559 Total reward: 67.28660003344217 Explore P: 0.0100
Episode: 562 Total reward: 3.406797250111893 Explore P: 0.0100
Episode: 564 Total reward: 220.1322979430358 Explore P: 0.0100
Episode: 566 Total reward: 236.61250889798006 Explore P: 0.0100
Episode: 568 Total reward: 15.731282601753872 Explore P: 0.0100
Episode: 571 Total reward: 187.85315079241994 Explore P: 0.0100
Episode: 573 Total reward: 335.99575483736936 Explore P: 0.0100
Episode: 576 Total reward: 111.72418196499348 Explore P: 0.0100
Episode: 577 Total reward: 181.10896042175597 Explore P: 0.0100
Episode: 579 Total reward: 171.1004178350171 Explore P: 0.0100
Episode: 580 Total reward: 325.9472937351287 Explore P: 0.0100
Episode: 582 Total reward: 144.10295623044163 Explore P: 0.0100
Episode: 584 Total reward: 16.611534873644512 Explore P: 0.0100
Episode: 586 Total reward: 39.01207189758618 Explore P: 0.0100
Episode: 588 Total reward: 53.047250598669045 Explore P: 0.0100
Episode: 590 Total reward: 5.2711451848348 Explore P: 0.0100
Episode: 591 Total reward: 297.1783389134408 Explore P: 0.0100
Episode: 593 Total reward: 15.7573069135348 Explore P: 0.0100
Episode: 594 Total reward: 319.7320291660725 Explore P: 0.0100
Episode: 596 Total reward: 32.419680198033646 Explore P: 0.0100
Episode: 598 Total reward: 24.88877775271732 Explore P: 0.0100
Episode: 599 Total reward: 185.106429250649 Explore P: 0.0100
Episode: 601 Total reward: 11.056602527697878 Explore P: 0.0100
Episode: 603 Total reward: 2.7565737366676313 Explore P: 0.0100
Episode: 604 Total reward: 139.1950407184801 Explore P: 0.0100
Episode: 606 Total reward: 29.712548901637398 Explore P: 0.0100
Episode: 607 Total reward: 274.4062169910567 Explore P: 0.0100
Episode: 608 Total reward: 133.65733752197903 Explore P: 0.0100


```
Episode: 609 Total reward: 162.18884913188722 Explore P: 0.0100
Episode: 610 Total reward: 266.01838727487376 Explore P: 0.0100
Episode: 612 Total reward: 18.78737422823907 Explore P: 0.0100
Episode: 614 Total reward: 4.942895432313283 Explore P: 0.0100
Episode: 615 Total reward: 300.2934268220656 Explore P: 0.0100
Episode: 617 Total reward: 37.60888212919236 Explore P: 0.0100
Episode: 618 Total reward: 319.1782815323773 Explore P: 0.0100
Episode: 619 Total reward: 280.61126141607093 Explore P: 0.0100
Episode: 620 Total reward: 308.1123381191478 Explore P: 0.0100
Episode: 621 Total reward: 333.53050872424245 Explore P: 0.0100
Episode: 626 Total reward: 228.64586252222426 Explore P: 0.0100
Episode: 668 Total reward: 31.033838133017184 Explore P: 0.0100
Episode: 669 Total reward: 10.751531168876681 Explore P: 0.0100
Episode: 670 Total reward: 11.806298950425116 Explore P: 0.0100
Episode: 671 Total reward: 6.88087114392935 Explore P: 0.0100
Episode: 672 Total reward: 8.75212601624662 Explore P: 0.0100
Episode: 673 Total reward: 9.677497791902473 Explore P: 0.0100
Episode: 674 Total reward: 7.742401716454576 Explore P: 0.0100
Episode: 675 Total reward: 9.78238147838662 Explore P: 0.0100
Episode: 676 Total reward: 7.686118437287708 Explore P: 0.0100
Episode: 677 Total reward: 9.720787685364485 Explore P: 0.0100
Episode: 678 Total reward: 9.828342661009325 Explore P: 0.0100
Episode: 679 Total reward: 9.81358415185241 Explore P: 0.0100
Episode: 680 Total reward: 7.66912721749395 Explore P: 0.0100
Episode: 681 Total reward: 6.906125693814828 Explore P: 0.0100
Episode: 682 Total reward: 9.838182206634277 Explore P: 0.0100
Episode: 683 Total reward: 7.702640296891333 Explore P: 0.0100
Episode: 684 Total reward: 8.887917550707547 Explore P: 0.0100
Episode: 685 Total reward: 10.78757845024423 Explore P: 0.0100
Episode: 686 Total reward: 10.770947302458808 Explore P: 0.0100
Episode: 687 Total reward: 8.59284636968126 Explore P: 0.0100
Episode: 688 Total reward: 8.7736664620849 Explore P: 0.0100
Episode: 689 Total reward: 7.858007690636441 Explore P: 0.0100
Episode: 690 Total reward: 6.838442519151917 Explore P: 0.0100
Episode: 691 Total reward: 10.771330296061933 Explore P: 0.0100
Episode: 692 Total reward: 10.747405609329387 Explore P: 0.0100
Episode: 693 Total reward: 7.777254668297246 Explore P: 0.0100
Episode: 694 Total reward: 8.792643379575262 Explore P: 0.0100
Episode: 695 Total reward: 6.915193799837045 Explore P: 0.0100
Episode: 696 Total reward: 9.840296093607321 Explore P: 0.0100
Episode: 697 Total reward: 9.83067769013966 Explore P: 0.0100
Episode: 698 Total reward: 8.699408107747635 Explore P: 0.0100
Episode: 699 Total reward: 7.6577400077755255 Explore P: 0.0100
```

Visualizing training

Below I plot the total rewards for each episode. The rolling average is plotted in blue.

```

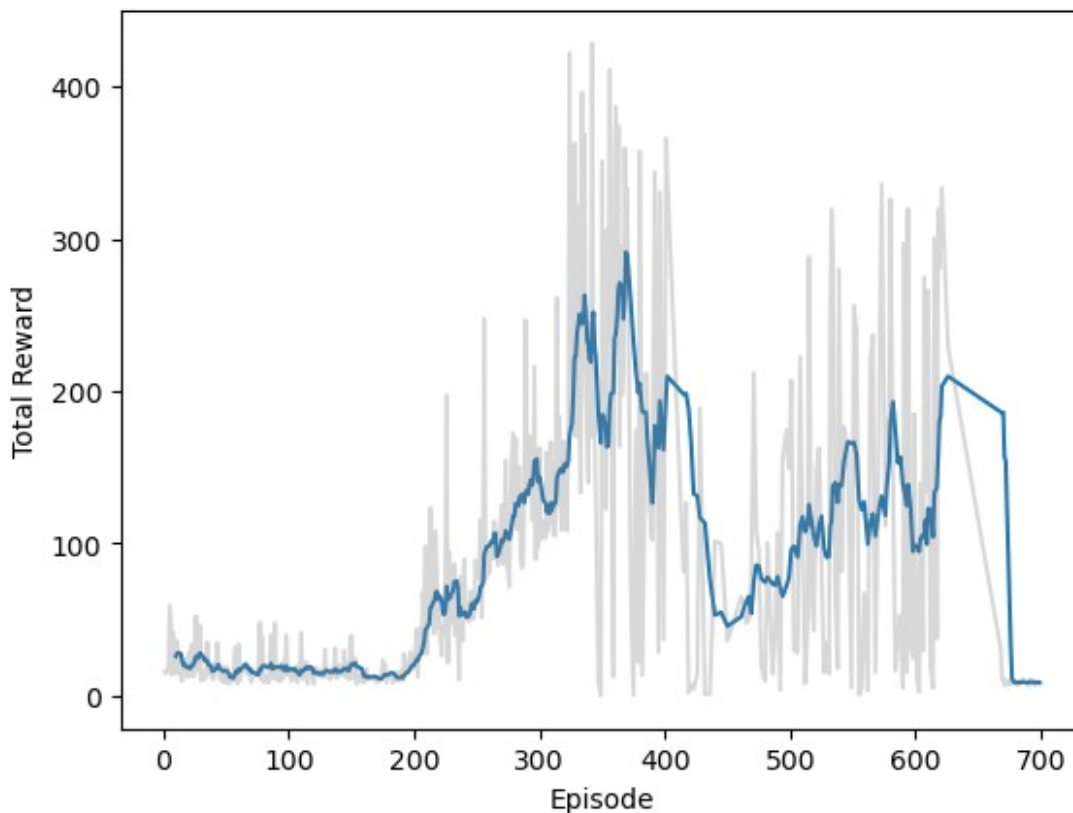
%matplotlib inline
import matplotlib.pyplot as plt

def running_mean(x, N):
    cumsum = np.cumsum(np.insert(x, 0, 0))
    return (cumsum[N:] - cumsum[:-N]) / N

eps, rews = np.array(rewards_list).T
smoothed_rews = running_mean(rews, 10)
plt.plot(eps[-len(smoothed_rews):], smoothed_rews)
plt.plot(eps, rews, color='grey', alpha=0.3)
plt.xlabel('Episode')
plt.ylabel('Total Reward')

Text(0, 0.5, 'Total Reward')

```



Watch a trained agent

Note that the episode ends after 500 steps.

```

def watch_agent(n_episodes):
    with tf.compat.v1.Session() as sess:
        saver.restore(sess,
"C:/Users/Shubham/Downloads/.ipynb_checkpoints/CartPole.ckpt")

```

```

    for episode in range(n_episodes):
        state_tuple = env.reset()
        r = 0
        while True:
            r += 1
            state = state_tuple[0]
            feed = {mainQN.inputs_: state.reshape((1,
*state.shape)))}
            Qs = sess.run(mainQN.output, feed_dict=feed)
            action = np.argmax(Qs)
            result = env.step(action)
            state, reward, done, _ = result[0], result[1],
result[2], result[3]
            env.render()
            sleep(0.05)
            if done:
                env.close()
                print(f"reward: {r}")
                break

```

watch_agent(3)

INFO:tensorflow:Restoring parameters from
C:/Users/Shubham/Downloads/.ipynb_checkpoints/CartPole.ckpt

C:\Users\Shubham\AppData\Roaming\Python\Python39\site-packages\gym\envs\classic_control\cartpole.py:211: UserWarning: WARN: You are calling render method without specifying any render mode. You can specify the render_mode at initialization, e.g. gym("CartPole-v1", render_mode="rgb_array")
gym.logger.warn(

reward: 9
reward: 10
reward: 9