LunarLanderDQN2025Assignment

March 10, 2025

0.1 Lunar lander with DQN-style neural function approximator using PyTorch

0.1.1 Christian Igel, 2025

If you have suggestions for improvement, let me know.

I took inspiration from https://github.com/udacity/deep-learning/blob/master/reinforcement/Q-learning-cart.ipynb.

Imports:

[25]:
```
import gymnasium as gym

from tqdm.notebook import tqdm  # Progress bar

import torch
import torch.nn as nn
import torch.nn.functional as F

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

Create the game environment (you need the `gym` package):

[26]:
```
env_visual = gym.make('LunarLander-v3', render_mode="human")
action_size = 4
state_size = 8
```

Let's just test the environment first:

[27]:
```
test_episodes = 0
for _ in range(test_episodes):
    R = 0
    state, _ = env_visual.reset()  # Environment starts in a random state, cart␣
 ↪and pole are moving
    print("initial state:", state)
    while True:  # Environment sets "truncated" to true after 500 steps
        env_visual.render()
        state, reward, terminated, truncated, _ = env_visual.step(env_visual.
 ↪action_space.sample()) #  Take a random action
```

```
        R += reward  # Accumulate reward
        if terminated or truncated:
            print("return: ", R)
            env_visual.reset()
            break
```

[28]:
```
#env.close()  # Closes the visualization window
```

Define Q network architecture:

[29]:
```
class QNetwork(nn.Module):
    def __init__(self, state_size=8, action_size=4, hidden_size=10, bias=True):
        super(QNetwork, self).__init__()
        self.fc1 = nn.Linear(state_size, hidden_size, bias)
        self.fc2 = nn.Linear(hidden_size, hidden_size, bias)
        self.output_layer = nn.Linear(hidden_size + state_size, action_size,␣
 ↪bias)

    def forward(self, x_input):
        x = F.tanh(self.fc1(x_input))
        x = F.tanh(self.fc2(x))
        x = torch.cat((x_input, x), dim=1)
        x = self.output_layer(x)
        return x
```

Data structure for storing experiences:

[30]:
```
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]
```

Define basic constants:

[31]:
```
train_episodes = 400          # Max number of episodes to learn from
gamma = 0.99                  # Future reward discount
learning_rate = 0.001         # Q-network learning rate
tau = .01                     # learning rate for target network

# Exploration parameters
```

```
explore_start = 1.0            # Exploration probability at start
explore_stop = 0.0001          # Minimum exploration probability
decay_rate = 0.05              # Exponential decay rate for exploration prob

# Network parameters
hidden_size = 64               # Number of units in each Q-network hidden layer

# Memory parameters
memory_size = 10000            # Memory capacity
batch_size = 128               # Experience mini-batch size
pretrain_length = batch_size   # Number experiences to pretrain the memory

log_path = "/tmp/deep_Q_network"
```

Instantiate network:

[32]:
```
mainQN = QNetwork(hidden_size=hidden_size)
print(mainQN)
```

```
QNetwork(
  (fc1): Linear(in_features=8, out_features=64, bias=True)
  (fc2): Linear(in_features=64, out_features=64, bias=True)
  (output_layer): Linear(in_features=72, out_features=4, bias=True)
)
```

Initialize the experience memory:

[33]:
```
# Initialize the simulation
env = gym.make('LunarLander-v3')
state = env.reset()[0]

memory = Memory(max_size=memory_size)

# Make a bunch of random actions and store the experiences
for _ in tqdm(range(pretrain_length)):
    # Make a random action
    action = env.action_space.sample()
    next_state, reward, terminated, truncated, _ = env.step(action)

    if terminated or truncated:
        # 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 new episode
        env.reset()
        # Take one random step to get the pole and cart moving
```

```
        state, reward, terminated, truncated, _ = env.step(env.action_space.
↪sample())
    else:
        # Add experience to memory
        memory.add((state, action, reward, next_state))
        state = next_state
```

```
  0%|          | 0/128 [00:00<?, ?it/s]
```

Now train with experiences:

[34]:
```
total_reward_list = []  # Returns for the individual episodes

optimizer = torch.optim.AdamW(mainQN.parameters(), lr=learning_rate) # AdamW␣
↪uses weight decay by default
loss_fn = torch.nn.MSELoss()

for ep in range(train_episodes):
    total_reward = 0  # Return / accumulated rewards
    state = env.reset()[0]  # Reset and get initial state
    while True:
        # Explore or exploit
        explore_p = explore_stop + (explore_start - explore_stop)*np.
↪exp(-decay_rate*ep)
        if explore_p > np.random.rand():
            # Pick a random action
            action = env.action_space.sample()
        else:
            # Get action from Q-network
            state_tensor = torch.from_numpy(np.resize(state, (1, state_size)).
↪astype(np.float32))
            Qs = mainQN(state_tensor)
            action = torch.argmax(Qs).item()

        # Take action, get new state and reward
        next_state, reward, terminated, truncated, _ = env.step(action)

        total_reward += reward  # Return / accumulated rewards

        if terminated or truncated:
            # Episode ends because of failure, so no next state
            next_state = np.zeros(state.shape)

            print('Episode: {}'.format(ep), 'Total reward: {}'.
↪format(total_reward),
                  'Training loss: {:.4f}'.format(loss), 'Explore P: {:.4f}'.
↪format(explore_p))
            total_reward_list.append((ep, total_reward))
```

```
            # Add experience to memory
            memory.add((state, action, reward, next_state))
            break; # End of episode
        else:
            # Add experience to memory
            memory.add((state, action, reward, next_state))
            state = next_state

        # Sample mini-batch from memory
        batch = memory.sample(batch_size)
        next_states_np = np.array([each[3] for each in batch], dtype=np.float32)
        next_states = torch.as_tensor(next_states_np)  # as_tensor does not␣
↪copy the data
        rewards     = torch.as_tensor(np.array([each[2] for each in batch],␣
↪dtype=np.float32))
        states      = torch.as_tensor(np.array([each[0] for each in batch],␣
↪dtype=np.float32))
        actions     = torch.as_tensor(np.array([each[1] for each in batch]))

        # Compute Q values for all actions in the new state
        target_Qs = mainQN(next_states)

        # Set target_Qs to 0 for states where episode ended because of failure
        episode_ends = (next_states_np == np.zeros(states[0].shape)).all(axis=1)
        target_Qs[episode_ends] = torch.zeros(action_size)

        # Compute targets
        max_elements = torch.max(target_Qs, dim=1)[0].detach()
        y = rewards + gamma * max_elements

        # Network learning starts here
        optimizer.zero_grad()

        # Compute the Q values of the actions taken
        main_Qs = mainQN(states)  # Q values for all action in each state
        Q = torch.gather(main_Qs, 1, actions.unsqueeze(-1)).squeeze()  # Only␣
↪the Q values for the actions taken

        # Gradient-based update
        loss = loss_fn(Q, y)
        loss.backward()
        optimizer.step()
```

```
Episode: 0 Total reward: -89.48895185144623 Training loss: 69.8484 Explore P:
1.0000
Episode: 1 Total reward: -206.7479823074944 Training loss: 61.8658 Explore P:
```

```
0.9512
Episode: 2 Total reward: -112.69879291469456 Training loss: 52.7202 Explore P:
0.9048
Episode: 3 Total reward: -38.590815418151124 Training loss: 107.3628 Explore P:
0.8607
Episode: 4 Total reward: -144.64825523135613 Training loss: 89.9198 Explore P:
0.8187
Episode: 5 Total reward: -119.10905674831314 Training loss: 10.2476 Explore P:
0.7788
Episode: 6 Total reward: -218.40983628087324 Training loss: 49.4395 Explore P:
0.7408
Episode: 7 Total reward: -165.38192796560642 Training loss: 4.4334 Explore P:
0.7047
Episode: 8 Total reward: -239.6680045370199 Training loss: 42.8873 Explore P:
0.6704
Episode: 9 Total reward: -215.7876083913913 Training loss: 34.2583 Explore P:
0.6377
Episode: 10 Total reward: -242.6327953350563 Training loss: 2.8893 Explore P:
0.6066
Episode: 11 Total reward: -480.24520581778665 Training loss: 74.5340 Explore P:
0.5770
Episode: 12 Total reward: -89.32082274054045 Training loss: 4.6947 Explore P:
0.5489
Episode: 13 Total reward: -433.3408251440758 Training loss: 6.7631 Explore P:
0.5221
Episode: 14 Total reward: -143.06458434477645 Training loss: 8.9373 Explore P:
0.4966
Episode: 15 Total reward: -107.46948805161945 Training loss: 51.9757 Explore P:
0.4724
Episode: 16 Total reward: -295.61234134023584 Training loss: 27.7782 Explore P:
0.4494
Episode: 17 Total reward: -7.864639761309533 Training loss: 46.0692 Explore P:
0.4275
Episode: 18 Total reward: -106.91317957729017 Training loss: 50.0501 Explore P:
0.4066
Episode: 19 Total reward: -308.8899691208443 Training loss: 45.3861 Explore P:
0.3868
Episode: 20 Total reward: -380.0962546017456 Training loss: 67.0883 Explore P:
0.3679
Episode: 21 Total reward: -108.242475864567 Training loss: 23.2723 Explore P:
0.3500
Episode: 22 Total reward: -235.8762891142914 Training loss: 23.6223 Explore P:
0.3329
Episode: 23 Total reward: -327.70129180003283 Training loss: 58.1270 Explore P:
0.3167
Episode: 24 Total reward: -126.68988244795915 Training loss: 32.6143 Explore P:
0.3013
Episode: 25 Total reward: -134.3829582485368 Training loss: 8.7045 Explore P:
```

```
0.2866
Episode: 26 Total reward: -157.83403448694142 Training loss: 36.0822 Explore P:
0.2726
Episode: 27 Total reward: -260.84326588824047 Training loss: 21.7956 Explore P:
0.2593
Episode: 28 Total reward: -156.25017984794272 Training loss: 31.8013 Explore P:
0.2467
Episode: 29 Total reward: -105.92122871826287 Training loss: 12.8848 Explore P:
0.2346
Episode: 30 Total reward: -90.23384514332787 Training loss: 17.2570 Explore P:
0.2232
Episode: 31 Total reward: -50.18930226398655 Training loss: 29.0710 Explore P:
0.2123
Episode: 32 Total reward: -223.52287432561673 Training loss: 51.5871 Explore P:
0.2020
Episode: 33 Total reward: -22.00710019947489 Training loss: 24.2114 Explore P:
0.1921
Episode: 34 Total reward: -362.99664796370894 Training loss: 15.9605 Explore P:
0.1828
Episode: 35 Total reward: -99.30926627919624 Training loss: 13.2295 Explore P:
0.1739
Episode: 36 Total reward: -115.0893650081561 Training loss: 13.4330 Explore P:
0.1654
Episode: 37 Total reward: 40.09348305951183 Training loss: 38.9293 Explore P:
0.1573
Episode: 38 Total reward: -142.36463013109397 Training loss: 74.4608 Explore P:
0.1497
Episode: 39 Total reward: -110.41465458432614 Training loss: 27.1938 Explore P:
0.1424
Episode: 40 Total reward: -88.24984834264112 Training loss: 9.2473 Explore P:
0.1354
Episode: 41 Total reward: -141.9872070642234 Training loss: 49.8156 Explore P:
0.1288
Episode: 42 Total reward: -137.94819544825896 Training loss: 14.0724 Explore P:
0.1225
Episode: 43 Total reward: -215.74755279807854 Training loss: 20.4029 Explore P:
0.1166
Episode: 44 Total reward: -45.55121678300274 Training loss: 3.1842 Explore P:
0.1109
Episode: 45 Total reward: -53.81604044873977 Training loss: 4.0919 Explore P:
0.1055
Episode: 46 Total reward: -65.37246381409436 Training loss: 2.1515 Explore P:
0.1003
Episode: 47 Total reward: -2.6988796931443324 Training loss: 8.5260 Explore P:
0.0955
Episode: 48 Total reward: 70.81205435773133 Training loss: 16.7507 Explore P:
0.0908
Episode: 49 Total reward: -27.718054832366064 Training loss: 2.5371 Explore P:
```

```
0.0864
Episode: 50 Total reward: -202.74837873099597 Training loss: 1.9928 Explore P:
0.0822
Episode: 51 Total reward: -195.384061306896 Training loss: 1.6368 Explore P:
0.0782
Episode: 52 Total reward: -84.11032187833783 Training loss: 2.1044 Explore P:
0.0744
Episode: 53 Total reward: 20.306257152458503 Training loss: 0.9103 Explore P:
0.0707
Episode: 54 Total reward: 150.57178481560624 Training loss: 1.3055 Explore P:
0.0673
Episode: 55 Total reward: -72.57679713833896 Training loss: 1.9497 Explore P:
0.0640
Episode: 56 Total reward: -51.39806479625563 Training loss: 1.1714 Explore P:
0.0609
Episode: 57 Total reward: -118.22320263579206 Training loss: 5.9674 Explore P:
0.0579
Episode: 58 Total reward: 51.33146740753294 Training loss: 1.0419 Explore P:
0.0551
Episode: 59 Total reward: -49.66936805328337 Training loss: 0.9781 Explore P:
0.0524
Episode: 60 Total reward: 135.7989554017953 Training loss: 8.8515 Explore P:
0.0499
Episode: 61 Total reward: -83.00176565856673 Training loss: 7.0910 Explore P:
0.0475
Episode: 62 Total reward: -393.3335271821846 Training loss: 1.8464 Explore P:
0.0451
Episode: 63 Total reward: -66.86362664369447 Training loss: 1.0874 Explore P:
0.0429
Episode: 64 Total reward: 2.637731419964208 Training loss: 1.0441 Explore P:
0.0409
Episode: 65 Total reward: -187.7388604201849 Training loss: 1.0370 Explore P:
0.0389
Episode: 66 Total reward: 103.19321899728429 Training loss: 1.6666 Explore P:
0.0370
Episode: 67 Total reward: 147.13666533175166 Training loss: 0.8479 Explore P:
0.0352
Episode: 68 Total reward: -90.66486972515875 Training loss: 1.2965 Explore P:
0.0335
Episode: 69 Total reward: 28.41375352027312 Training loss: 0.7749 Explore P:
0.0318
Episode: 70 Total reward: 114.44581792070466 Training loss: 1.1662 Explore P:
0.0303
Episode: 71 Total reward: 203.7252791238643 Training loss: 10.1486 Explore P:
0.0288
Episode: 72 Total reward: -92.86862688495847 Training loss: 1.0240 Explore P:
0.0274
Episode: 73 Total reward: 126.71269418993569 Training loss: 0.9901 Explore P:
```

```
0.0261
Episode: 74 Total reward: 228.95779956377686 Training loss: 1.0566 Explore P:
0.0248
Episode: 75 Total reward: 162.29433561792095 Training loss: 1.2065 Explore P:
0.0236
Episode: 76 Total reward: 144.94623425010286 Training loss: 12.7214 Explore P:
0.0225
Episode: 77 Total reward: -26.636636349770058 Training loss: 1.1044 Explore P:
0.0214
Episode: 78 Total reward: -236.05741569180375 Training loss: 1.0837 Explore P:
0.0203
Episode: 79 Total reward: -107.23743798275 Training loss: 1.1371 Explore P:
0.0194
Episode: 80 Total reward: -502.4806986698471 Training loss: 5.7747 Explore P:
0.0184
Episode: 81 Total reward: 225.1482609301147 Training loss: 11.0987 Explore P:
0.0175
Episode: 82 Total reward: 113.91136247347936 Training loss: 0.9283 Explore P:
0.0167
Episode: 83 Total reward: -252.45928159199008 Training loss: 0.9223 Explore P:
0.0159
Episode: 84 Total reward: 51.165715120313706 Training loss: 0.9824 Explore P:
0.0151
Episode: 85 Total reward: -69.02434187396659 Training loss: 3.2469 Explore P:
0.0144
Episode: 86 Total reward: 33.475805118790376 Training loss: 6.0320 Explore P:
0.0137
Episode: 87 Total reward: -15.554729385183016 Training loss: 2.4293 Explore P:
0.0130
Episode: 88 Total reward: 211.09538840647085 Training loss: 2.5711 Explore P:
0.0124
Episode: 89 Total reward: 177.65716362926017 Training loss: 3.4372 Explore P:
0.0118
Episode: 90 Total reward: 28.9615644930313 Training loss: 9.6361 Explore P:
0.0112
Episode: 91 Total reward: 30.885836527152986 Training loss: 0.9598 Explore P:
0.0107
Episode: 92 Total reward: 123.05962115939406 Training loss: 1.4930 Explore P:
0.0102
Episode: 93 Total reward: 233.5758385099452 Training loss: 1.0564 Explore P:
0.0097
Episode: 94 Total reward: 231.70195544077862 Training loss: 0.9707 Explore P:
0.0092
Episode: 95 Total reward: 238.05399655622446 Training loss: 2.6969 Explore P:
0.0088
Episode: 96 Total reward: 228.86732637634722 Training loss: 0.6633 Explore P:
0.0083
Episode: 97 Total reward: 226.47519801212792 Training loss: 1.3545 Explore P:
```

```
0.0079
Episode: 98 Total reward: 243.07349220013973 Training loss: 3.1945 Explore P:
0.0075
Episode: 99 Total reward: 212.27511245756511 Training loss: 1.3791 Explore P:
0.0072
Episode: 100 Total reward: 253.62019944713944 Training loss: 0.9471 Explore P:
0.0068
Episode: 101 Total reward: 253.4118647814138 Training loss: 0.6430 Explore P:
0.0065
Episode: 102 Total reward: 54.66378767092445 Training loss: 0.6831 Explore P:
0.0062
Episode: 103 Total reward: -51.2370638259805 Training loss: 0.8039 Explore P:
0.0059
Episode: 104 Total reward: -160.29183454979312 Training loss: 10.9328 Explore P:
0.0056
Episode: 105 Total reward: 227.35912602082277 Training loss: 1.6847 Explore P:
0.0053
Episode: 106 Total reward: 184.97048110118791 Training loss: 0.9204 Explore P:
0.0051
Episode: 107 Total reward: 205.51460976148877 Training loss: 1.1800 Explore P:
0.0048
Episode: 108 Total reward: 193.97781906258749 Training loss: 1.5501 Explore P:
0.0046
Episode: 109 Total reward: 232.2305328349691 Training loss: 1.8077 Explore P:
0.0044
Episode: 110 Total reward: 236.13179093015003 Training loss: 1.4823 Explore P:
0.0042
Episode: 111 Total reward: 254.11363244676426 Training loss: 1.4943 Explore P:
0.0040
Episode: 112 Total reward: 258.01689942039457 Training loss: 1.4507 Explore P:
0.0038
Episode: 113 Total reward: 245.64511421794887 Training loss: 1.2711 Explore P:
0.0036
Episode: 114 Total reward: 250.31789187783943 Training loss: 0.9635 Explore P:
0.0034
Episode: 115 Total reward: 230.36090588923975 Training loss: 1.4349 Explore P:
0.0033
Episode: 116 Total reward: 261.8585253780519 Training loss: 0.9805 Explore P:
0.0031
Episode: 117 Total reward: -141.65907801339645 Training loss: 182.2961 Explore
P: 0.0030
Episode: 118 Total reward: 191.8138836002774 Training loss: 1.6104 Explore P:
0.0028
Episode: 119 Total reward: 235.73849807322247 Training loss: 1.0691 Explore P:
0.0027
Episode: 120 Total reward: -11.411382107789947 Training loss: 8.0292 Explore P:
0.0026
Episode: 121 Total reward: -93.02299862239482 Training loss: 1.4774 Explore P:
```

```
0.0025
Episode: 122 Total reward: 165.1902752596016 Training loss: 1.0479 Explore P:
0.0023
Episode: 123 Total reward: 78.46333030997214 Training loss: 4.6233 Explore P:
0.0022
Episode: 124 Total reward: 152.70983662877063 Training loss: 2.6541 Explore P:
0.0021
Episode: 125 Total reward: 261.42017238139687 Training loss: 1.0385 Explore P:
0.0020
Episode: 126 Total reward: 218.48202520695673 Training loss: 2.4148 Explore P:
0.0019
Episode: 127 Total reward: 235.5319093618593 Training loss: 1.4651 Explore P:
0.0018
Episode: 128 Total reward: 228.5895401491494 Training loss: 1.7683 Explore P:
0.0018
Episode: 129 Total reward: 176.9558719900173 Training loss: 1.2060 Explore P:
0.0017
Episode: 130 Total reward: 171.3044840530344 Training loss: 8.3725 Explore P:
0.0016
Episode: 131 Total reward: 175.17089806411025 Training loss: 1.0007 Explore P:
0.0015
Episode: 132 Total reward: 218.98479371909434 Training loss: 1.9759 Explore P:
0.0015
Episode: 133 Total reward: -7.669204483518612 Training loss: 8.5777 Explore P:
0.0014
Episode: 134 Total reward: 146.53705538098032 Training loss: 1.0915 Explore P:
0.0013
Episode: 135 Total reward: 223.32244046719455 Training loss: 0.8772 Explore P:
0.0013
Episode: 136 Total reward: 247.5670425710722 Training loss: 1.4812 Explore P:
0.0012
Episode: 137 Total reward: 206.13352956047058 Training loss: 1.5507 Explore P:
0.0012
Episode: 138 Total reward: 237.23946197343002 Training loss: 21.5041 Explore P:
0.0011
Episode: 139 Total reward: 233.8796388807231 Training loss: 1.0854 Explore P:
0.0011
Episode: 140 Total reward: 160.65003587868787 Training loss: 2.0207 Explore P:
0.0010
Episode: 141 Total reward: -66.21986361529338 Training loss: 4.2154 Explore P:
0.0010
Episode: 142 Total reward: -235.67059711713597 Training loss: 2.7536 Explore P:
0.0009
Episode: 143 Total reward: -15.536594653595017 Training loss: 1.7572 Explore P:
0.0009
Episode: 144 Total reward: -60.64526080104163 Training loss: 4.0121 Explore P:
0.0008
Episode: 145 Total reward: 220.10642832330865 Training loss: 9.0504 Explore P:
```

```
0.0008
Episode: 146 Total reward: 277.2128519062611 Training loss: 3.7179 Explore P:
0.0008
Episode: 147 Total reward: -67.4704511654564 Training loss: 2.7601 Explore P:
0.0007
Episode: 148 Total reward: 264.41397184886057 Training loss: 12.0218 Explore P:
0.0007
Episode: 149 Total reward: 186.91014946469878 Training loss: 1.1151 Explore P:
0.0007
Episode: 150 Total reward: 240.6531219423966 Training loss: 18.6067 Explore P:
0.0007
Episode: 151 Total reward: 207.87843096252033 Training loss: 1.3258 Explore P:
0.0006
Episode: 152 Total reward: 264.2455875113509 Training loss: 1.5352 Explore P:
0.0006
Episode: 153 Total reward: 206.91680102994485 Training loss: 1.3189 Explore P:
0.0006
Episode: 154 Total reward: 223.41968745229843 Training loss: 2.7007 Explore P:
0.0006
Episode: 155 Total reward: 229.56234212085428 Training loss: 1.4410 Explore P:
0.0005
Episode: 156 Total reward: 212.18193199462178 Training loss: 2.3670 Explore P:
0.0005
Episode: 157 Total reward: -498.5288106081505 Training loss: 3.3914 Explore P:
0.0005
Episode: 158 Total reward: -263.12015611415137 Training loss: 1.7614 Explore P:
0.0005
Episode: 159 Total reward: -208.43875102447572 Training loss: 96.5537 Explore P:
0.0005
Episode: 160 Total reward: 212.56070429443037 Training loss: 7.1517 Explore P:
0.0004
Episode: 161 Total reward: 214.56436843338724 Training loss: 2.1804 Explore P:
0.0004
Episode: 162 Total reward: 203.90610050993865 Training loss: 12.5669 Explore P:
0.0004
Episode: 163 Total reward: 243.61276406799513 Training loss: 72.1409 Explore P:
0.0004
Episode: 164 Total reward: 262.9910553696503 Training loss: 1.4685 Explore P:
0.0004
Episode: 165 Total reward: 210.76992919997588 Training loss: 1.3550 Explore P:
0.0004
Episode: 166 Total reward: 165.50477240830753 Training loss: 1.2309 Explore P:
0.0003
Episode: 167 Total reward: -137.6382694710179 Training loss: 1.9468 Explore P:
0.0003
Episode: 168 Total reward: 198.6353224273828 Training loss: 1.7975 Explore P:
0.0003
Episode: 169 Total reward: 215.10206696437587 Training loss: 4.1170 Explore P:
```

```
0.0003
Episode: 170 Total reward: 291.0712975257157 Training loss: 1.1410 Explore P:
0.0003
Episode: 171 Total reward: 31.246240840250067 Training loss: 1.2099 Explore P:
0.0003
Episode: 172 Total reward: 171.13108658476244 Training loss: 12.7350 Explore P:
0.0003
Episode: 173 Total reward: 265.4536168129573 Training loss: 4.2515 Explore P:
0.0003
Episode: 174 Total reward: 225.1276086336332 Training loss: 1.4898 Explore P:
0.0003
Episode: 175 Total reward: 258.1895544985696 Training loss: 4.7229 Explore P:
0.0003
Episode: 176 Total reward: 258.2520752170392 Training loss: 21.8969 Explore P:
0.0003
Episode: 177 Total reward: 238.19835186272334 Training loss: 0.9617 Explore P:
0.0002
Episode: 178 Total reward: -69.08234354930552 Training loss: 4.4072 Explore P:
0.0002
Episode: 179 Total reward: -35.50790074679162 Training loss: 39.3401 Explore P:
0.0002
Episode: 180 Total reward: 222.3363121007107 Training loss: 2.0433 Explore P:
0.0002
Episode: 181 Total reward: 199.09624964721448 Training loss: 5.3491 Explore P:
0.0002
Episode: 182 Total reward: 119.8807269969169 Training loss: 1.2268 Explore P:
0.0002
Episode: 183 Total reward: 250.64247455845634 Training loss: 2.3849 Explore P:
0.0002
Episode: 184 Total reward: 270.7430103122955 Training loss: 2.1484 Explore P:
0.0002
Episode: 185 Total reward: 233.00527335249814 Training loss: 2.4311 Explore P:
0.0002
Episode: 186 Total reward: 237.37717753593412 Training loss: 2.1666 Explore P:
0.0002
Episode: 187 Total reward: 196.67969151574357 Training loss: 3.0221 Explore P:
0.0002
Episode: 188 Total reward: 204.45535206199037 Training loss: 3.1449 Explore P:
0.0002
Episode: 189 Total reward: -196.68282495904225 Training loss: 4.2565 Explore P:
0.0002
Episode: 190 Total reward: 79.3648876696916 Training loss: 4.1734 Explore P:
0.0002
Episode: 191 Total reward: 227.01722590361538 Training loss: 2.5838 Explore P:
0.0002
Episode: 192 Total reward: 213.43774854607722 Training loss: 2.3124 Explore P:
0.0002
Episode: 193 Total reward: 206.73063837626393 Training loss: 1.3068 Explore P:
```

```
0.0002
Episode: 194 Total reward: 198.22613009359222 Training loss: 8.8344 Explore P:
0.0002
Episode: 195 Total reward: 259.1952323504589 Training loss: 3.6123 Explore P:
0.0002
Episode: 196 Total reward: 213.1710698433467 Training loss: 1.0074 Explore P:
0.0002
Episode: 197 Total reward: 233.81583796946424 Training loss: 3.2964 Explore P:
0.0002
Episode: 198 Total reward: 259.52680461757484 Training loss: 11.4526 Explore P:
0.0002
Episode: 199 Total reward: 219.19366814359324 Training loss: 2.5717 Explore P:
0.0001
Episode: 200 Total reward: 223.20855833869817 Training loss: 2.7582 Explore P:
0.0001
Episode: 201 Total reward: 241.97047215126076 Training loss: 3.3939 Explore P:
0.0001
Episode: 202 Total reward: 254.1847359013308 Training loss: 1.7421 Explore P:
0.0001
Episode: 203 Total reward: 228.76705623361 Training loss: 3.8959 Explore P:
0.0001
Episode: 204 Total reward: 236.16521504082527 Training loss: 1.6704 Explore P:
0.0001
Episode: 205 Total reward: -62.306758407553964 Training loss: 15.0124 Explore P:
0.0001
Episode: 206 Total reward: 222.22392858642263 Training loss: 1.2905 Explore P:
0.0001
Episode: 207 Total reward: 294.18149426322054 Training loss: 0.9844 Explore P:
0.0001
Episode: 208 Total reward: 249.62450465515536 Training loss: 8.8104 Explore P:
0.0001
Episode: 209 Total reward: 270.4930129635949 Training loss: 17.8313 Explore P:
0.0001
Episode: 210 Total reward: 217.060453502791 Training loss: 1.8428 Explore P:
0.0001
Episode: 211 Total reward: 244.96988662889007 Training loss: 2.0075 Explore P:
0.0001
Episode: 212 Total reward: 234.8522793997102 Training loss: 1.3781 Explore P:
0.0001
Episode: 213 Total reward: 267.2111077868292 Training loss: 1.3747 Explore P:
0.0001
Episode: 214 Total reward: 259.20295949060073 Training loss: 1.1796 Explore P:
0.0001
Episode: 215 Total reward: 274.59173401743465 Training loss: 1.4382 Explore P:
0.0001
Episode: 216 Total reward: 293.6913187583184 Training loss: 5.5085 Explore P:
0.0001
Episode: 217 Total reward: 11.549461210920299 Training loss: 0.9089 Explore P:
```

```
0.0001
Episode: 218 Total reward: 256.97844105837487 Training loss: 0.7228 Explore P:
0.0001
Episode: 219 Total reward: 29.261309911785986 Training loss: 10.6444 Explore P:
0.0001
Episode: 220 Total reward: 257.4015914965671 Training loss: 0.8391 Explore P:
0.0001
Episode: 221 Total reward: 231.63091062757832 Training loss: 10.1117 Explore P:
0.0001
Episode: 222 Total reward: 257.57904011199594 Training loss: 1.8415 Explore P:
0.0001
Episode: 223 Total reward: 256.08812380455356 Training loss: 2.2741 Explore P:
0.0001
Episode: 224 Total reward: 256.91703475718174 Training loss: 2.3002 Explore P:
0.0001
Episode: 225 Total reward: 243.11576038999902 Training loss: 9.8383 Explore P:
0.0001
Episode: 226 Total reward: 245.76779107012152 Training loss: 115.8104 Explore P:
0.0001
Episode: 227 Total reward: -199.07394266162999 Training loss: 3.0827 Explore P:
0.0001
Episode: 228 Total reward: 217.4259666005843 Training loss: 2.7419 Explore P:
0.0001
Episode: 229 Total reward: -4.415707467835347 Training loss: 1.1872 Explore P:
0.0001
Episode: 230 Total reward: 219.51534851719248 Training loss: 1.0907 Explore P:
0.0001
Episode: 231 Total reward: 264.100143497048 Training loss: 2.9913 Explore P:
0.0001
Episode: 232 Total reward: 250.56599081735737 Training loss: 2.4990 Explore P:
0.0001
Episode: 233 Total reward: 265.2586590492917 Training loss: 3.3853 Explore P:
0.0001
Episode: 234 Total reward: 245.25685409990865 Training loss: 6.3139 Explore P:
0.0001
Episode: 235 Total reward: -157.93862315153828 Training loss: 4.2916 Explore P:
0.0001
Episode: 236 Total reward: 243.15503081793761 Training loss: 1.9643 Explore P:
0.0001
Episode: 237 Total reward: 148.64083681342612 Training loss: 1.4176 Explore P:
0.0001
Episode: 238 Total reward: 266.00717258698495 Training loss: 1.4516 Explore P:
0.0001
Episode: 239 Total reward: 261.3087459667945 Training loss: 1.6604 Explore P:
0.0001
Episode: 240 Total reward: 255.99508631611923 Training loss: 1.1678 Explore P:
0.0001
Episode: 241 Total reward: -203.13193320640266 Training loss: 2.2902 Explore P:
```

```
0.0001
Episode: 242 Total reward: 220.81068653946062 Training loss: 1.2341 Explore P:
0.0001
Episode: 243 Total reward: 256.00433566103425 Training loss: 1.1153 Explore P:
0.0001
Episode: 244 Total reward: 253.52108353510508 Training loss: 1.0503 Explore P:
0.0001
Episode: 245 Total reward: 216.2316315949958 Training loss: 4.8340 Explore P:
0.0001
Episode: 246 Total reward: 260.18170160182416 Training loss: 4.7259 Explore P:
0.0001
Episode: 247 Total reward: 229.6023921700572 Training loss: 2.1198 Explore P:
0.0001
Episode: 248 Total reward: 245.0911315883263 Training loss: 1.4580 Explore P:
0.0001
Episode: 249 Total reward: 118.41038454245631 Training loss: 1.5487 Explore P:
0.0001
Episode: 250 Total reward: 30.46310176902881 Training loss: 5.8746 Explore P:
0.0001
Episode: 251 Total reward: 248.27270921461553 Training loss: 1.3422 Explore P:
0.0001
Episode: 252 Total reward: 249.65197441734577 Training loss: 7.1865 Explore P:
0.0001
Episode: 253 Total reward: 234.03119092020123 Training loss: 2.0639 Explore P:
0.0001
Episode: 254 Total reward: 246.5687464745335 Training loss: 2.5651 Explore P:
0.0001
Episode: 255 Total reward: 252.87443838068148 Training loss: 1.1973 Explore P:
0.0001
Episode: 256 Total reward: 228.3237673204946 Training loss: 144.5600 Explore P:
0.0001
Episode: 257 Total reward: 238.65556112982742 Training loss: 1.0696 Explore P:
0.0001
Episode: 258 Total reward: 265.4196127376603 Training loss: 1.4849 Explore P:
0.0001
Episode: 259 Total reward: 273.594350475464 Training loss: 0.6683 Explore P:
0.0001
Episode: 260 Total reward: 11.138519631145215 Training loss: 0.9332 Explore P:
0.0001
Episode: 261 Total reward: 261.88594905382075 Training loss: 2.8748 Explore P:
0.0001
Episode: 262 Total reward: 248.93030895039288 Training loss: 14.2057 Explore P:
0.0001
Episode: 263 Total reward: 263.8809936817825 Training loss: 1.3803 Explore P:
0.0001
Episode: 264 Total reward: 273.5180091052265 Training loss: 2.1431 Explore P:
0.0001
Episode: 265 Total reward: 267.47271635198217 Training loss: 1.1363 Explore P:
```

```
0.0001
Episode: 266 Total reward: 224.6674553402695 Training loss: 1.9209 Explore P:
0.0001
Episode: 267 Total reward: 249.94040134805542 Training loss: 1.0522 Explore P:
0.0001
Episode: 268 Total reward: 241.99657602208796 Training loss: 0.7045 Explore P:
0.0001
Episode: 269 Total reward: 261.2499911712273 Training loss: 2.1626 Explore P:
0.0001
Episode: 270 Total reward: 268.95810746507607 Training loss: 0.9695 Explore P:
0.0001
Episode: 271 Total reward: 246.97712567376897 Training loss: 0.4245 Explore P:
0.0001
Episode: 272 Total reward: 249.31272607459846 Training loss: 0.7411 Explore P:
0.0001
Episode: 273 Total reward: 260.86202335837936 Training loss: 2.2592 Explore P:
0.0001
Episode: 274 Total reward: 266.46149896338903 Training loss: 8.1245 Explore P:
0.0001
Episode: 275 Total reward: 243.89002172806283 Training loss: 0.6912 Explore P:
0.0001
Episode: 276 Total reward: 251.77214119555467 Training loss: 7.3725 Explore P:
0.0001
Episode: 277 Total reward: 268.22049113324 Training loss: 0.9097 Explore P:
0.0001
Episode: 278 Total reward: 255.03548029883825 Training loss: 7.8011 Explore P:
0.0001
Episode: 279 Total reward: 223.12522489849368 Training loss: 0.8039 Explore P:
0.0001
Episode: 280 Total reward: 261.4267459991985 Training loss: 2.1637 Explore P:
0.0001
Episode: 281 Total reward: 257.44125697788473 Training loss: 1.8669 Explore P:
0.0001
Episode: 282 Total reward: 245.50366529663808 Training loss: 4.9866 Explore P:
0.0001
Episode: 283 Total reward: 245.16814847669522 Training loss: 5.6516 Explore P:
0.0001
Episode: 284 Total reward: 251.0933731639306 Training loss: 1.2868 Explore P:
0.0001
Episode: 285 Total reward: 239.668511240775 Training loss: 0.6746 Explore P:
0.0001
Episode: 286 Total reward: 215.03675106626423 Training loss: 3.4254 Explore P:
0.0001
Episode: 287 Total reward: 260.7177565781328 Training loss: 0.6782 Explore P:
0.0001
Episode: 288 Total reward: 272.1544371220565 Training loss: 1.1289 Explore P:
0.0001
Episode: 289 Total reward: 240.2618715668195 Training loss: 1.3876 Explore P:
```

```
0.0001
Episode: 290 Total reward: 259.03913910400934 Training loss: 6.0877 Explore P:
0.0001
Episode: 291 Total reward: 262.1444154423698 Training loss: 1.2551 Explore P:
0.0001
Episode: 292 Total reward: 197.74926966247028 Training loss: 1.8987 Explore P:
0.0001
Episode: 293 Total reward: 269.5987924888471 Training loss: 1.2576 Explore P:
0.0001
Episode: 294 Total reward: 237.47438804154993 Training loss: 0.9134 Explore P:
0.0001
Episode: 295 Total reward: 248.68533712271838 Training loss: 0.8375 Explore P:
0.0001
Episode: 296 Total reward: 276.25741279591006 Training loss: 0.6511 Explore P:
0.0001
Episode: 297 Total reward: 253.88197400403737 Training loss: 4.6477 Explore P:
0.0001
Episode: 298 Total reward: 275.9321939295865 Training loss: 0.8103 Explore P:
0.0001
Episode: 299 Total reward: 14.657135516641603 Training loss: 3.4260 Explore P:
0.0001
Episode: 300 Total reward: 265.7353520856482 Training loss: 0.9124 Explore P:
0.0001
Episode: 301 Total reward: 258.9959966229342 Training loss: 0.9835 Explore P:
0.0001
Episode: 302 Total reward: 285.14270204908655 Training loss: 3.6498 Explore P:
0.0001
Episode: 303 Total reward: 224.50920049455667 Training loss: 1.2490 Explore P:
0.0001
Episode: 304 Total reward: 1.0953206506708 Training loss: 0.8623 Explore P:
0.0001
Episode: 305 Total reward: 252.40690458634785 Training loss: 1.4532 Explore P:
0.0001
Episode: 306 Total reward: 256.5935449616669 Training loss: 5.6751 Explore P:
0.0001
Episode: 307 Total reward: 28.53141685371071 Training loss: 1.8455 Explore P:
0.0001
Episode: 308 Total reward: 274.7242203045371 Training loss: 1.8614 Explore P:
0.0001
Episode: 309 Total reward: 268.8249093008191 Training loss: 8.2638 Explore P:
0.0001
Episode: 310 Total reward: -511.9685624605492 Training loss: 5.4150 Explore P:
0.0001
Episode: 311 Total reward: -260.192577767456 Training loss: 3.2110 Explore P:
0.0001
Episode: 312 Total reward: 269.32667094349915 Training loss: 6.7365 Explore P:
0.0001
Episode: 313 Total reward: 227.8587871831914 Training loss: 2.0999 Explore P:
```

```
0.0001
Episode: 314 Total reward: 199.10888965597596 Training loss: 1.0764 Explore P:
0.0001
Episode: 315 Total reward: 236.24384278203195 Training loss: 6.7380 Explore P:
0.0001
Episode: 316 Total reward: 257.8313018693966 Training loss: 1.4009 Explore P:
0.0001
Episode: 317 Total reward: -158.19842890013223 Training loss: 1.4560 Explore P:
0.0001
Episode: 318 Total reward: 152.88852971836184 Training loss: 5.3351 Explore P:
0.0001
Episode: 319 Total reward: -262.2310500567946 Training loss: 6.6844 Explore P:
0.0001
Episode: 320 Total reward: 178.46736454166427 Training loss: 5.5511 Explore P:
0.0001
Episode: 321 Total reward: 3.066185098215925 Training loss: 2.6729 Explore P:
0.0001
Episode: 322 Total reward: -205.46303361803274 Training loss: 5.8042 Explore P:
0.0001
Episode: 323 Total reward: 208.48268899435533 Training loss: 4.6732 Explore P:
0.0001
Episode: 324 Total reward: 174.79363257761094 Training loss: 162.8316 Explore P:
0.0001
Episode: 325 Total reward: 235.64806106609086 Training loss: 11.4525 Explore P:
0.0001
Episode: 326 Total reward: 278.66747487301257 Training loss: 2.2771 Explore P:
0.0001
Episode: 327 Total reward: 230.30621169601838 Training loss: 1.5579 Explore P:
0.0001
Episode: 328 Total reward: 252.0702131110179 Training loss: 4.7827 Explore P:
0.0001
Episode: 329 Total reward: 222.48263556012947 Training loss: 7.8065 Explore P:
0.0001
Episode: 330 Total reward: 270.6728517555891 Training loss: 1.9841 Explore P:
0.0001
Episode: 331 Total reward: 270.53647898829206 Training loss: 3.0734 Explore P:
0.0001
Episode: 332 Total reward: 256.1856332709526 Training loss: 42.7864 Explore P:
0.0001
Episode: 333 Total reward: 279.01248229083683 Training loss: 4.3283 Explore P:
0.0001
Episode: 334 Total reward: -9.040242393412697 Training loss: 1.0085 Explore P:
0.0001
Episode: 335 Total reward: 261.15739398870915 Training loss: 4.9258 Explore P:
0.0001
Episode: 336 Total reward: 204.09163298205993 Training loss: 2.5711 Explore P:
0.0001
Episode: 337 Total reward: -157.1541727120039 Training loss: 9.4769 Explore P:
```

```
0.0001
Episode: 338 Total reward: 268.9807148020582 Training loss: 5.5701 Explore P:
0.0001
Episode: 339 Total reward: 209.12960738141953 Training loss: 4.5029 Explore P:
0.0001
Episode: 340 Total reward: 230.13884393971557 Training loss: 2.5278 Explore P:
0.0001
Episode: 341 Total reward: 240.54490415007768 Training loss: 1.7855 Explore P:
0.0001
Episode: 342 Total reward: 229.0990330040169 Training loss: 5.2180 Explore P:
0.0001
Episode: 343 Total reward: 271.21301742397395 Training loss: 23.7436 Explore P:
0.0001
Episode: 344 Total reward: 276.9323693071514 Training loss: 9.4295 Explore P:
0.0001
Episode: 345 Total reward: 265.6643934853491 Training loss: 2.8720 Explore P:
0.0001
Episode: 346 Total reward: 122.31262035734467 Training loss: 2.7743 Explore P:
0.0001
Episode: 347 Total reward: 194.72889768729678 Training loss: 5.7817 Explore P:
0.0001
Episode: 348 Total reward: 32.37788013070991 Training loss: 1.3983 Explore P:
0.0001
Episode: 349 Total reward: 285.0003478251709 Training loss: 4.4648 Explore P:
0.0001
Episode: 350 Total reward: 209.99730512882047 Training loss: 1.4921 Explore P:
0.0001
Episode: 351 Total reward: 272.7318585658551 Training loss: 1.7471 Explore P:
0.0001
Episode: 352 Total reward: 235.26063385520519 Training loss: 1.2467 Explore P:
0.0001
Episode: 353 Total reward: 215.72141838151802 Training loss: 2.3750 Explore P:
0.0001
Episode: 354 Total reward: 245.31797217909286 Training loss: 1.8048 Explore P:
0.0001
Episode: 355 Total reward: 288.6344279726086 Training loss: 10.8322 Explore P:
0.0001
Episode: 356 Total reward: 256.8109247687429 Training loss: 0.9814 Explore P:
0.0001
Episode: 357 Total reward: -148.11800454253586 Training loss: 1.3423 Explore P:
0.0001
Episode: 358 Total reward: 234.66485484285812 Training loss: 1.8405 Explore P:
0.0001
Episode: 359 Total reward: -38.33237355424709 Training loss: 1.1797 Explore P:
0.0001
Episode: 360 Total reward: 216.9605214950332 Training loss: 1.8517 Explore P:
0.0001
Episode: 361 Total reward: 244.95829003257518 Training loss: 1.6129 Explore P:
```

```
0.0001
Episode: 362 Total reward: -406.219039404381 Training loss: 4.3979 Explore P:
0.0001
Episode: 363 Total reward: -12.610108872057936 Training loss: 18.2727 Explore P:
0.0001
Episode: 364 Total reward: 138.89523641648 Training loss: 3.0486 Explore P:
0.0001
Episode: 365 Total reward: 231.3020638549704 Training loss: 10.1114 Explore P:
0.0001
Episode: 366 Total reward: 277.0040070284617 Training loss: 2.7316 Explore P:
0.0001
Episode: 367 Total reward: 242.18771772240098 Training loss: 2.5512 Explore P:
0.0001
Episode: 368 Total reward: 260.48483871249783 Training loss: 21.5002 Explore P:
0.0001
Episode: 369 Total reward: 200.3589555558703 Training loss: 1.8001 Explore P:
0.0001
Episode: 370 Total reward: 255.89409617422461 Training loss: 1.6227 Explore P:
0.0001
Episode: 371 Total reward: 256.0459344541758 Training loss: 2.1369 Explore P:
0.0001
Episode: 372 Total reward: 254.04227787787332 Training loss: 1.6005 Explore P:
0.0001
Episode: 373 Total reward: 83.97145384680827 Training loss: 1.4023 Explore P:
0.0001
Episode: 374 Total reward: 207.2629441614373 Training loss: 6.7873 Explore P:
0.0001
Episode: 375 Total reward: 149.3685358170792 Training loss: 22.1301 Explore P:
0.0001
Episode: 376 Total reward: 261.7870812602164 Training loss: 4.9908 Explore P:
0.0001
Episode: 377 Total reward: 254.86913957875154 Training loss: 0.6896 Explore P:
0.0001
Episode: 378 Total reward: 258.34681473341243 Training loss: 1.5848 Explore P:
0.0001
Episode: 379 Total reward: -362.50470412900137 Training loss: 1.4122 Explore P:
0.0001
Episode: 380 Total reward: 219.15330831198753 Training loss: 1.2317 Explore P:
0.0001
Episode: 381 Total reward: -20.90899393603901 Training loss: 8.7449 Explore P:
0.0001
Episode: 382 Total reward: 226.64503668406064 Training loss: 7.9784 Explore P:
0.0001
Episode: 383 Total reward: 253.73374964444284 Training loss: 1.5132 Explore P:
0.0001
Episode: 384 Total reward: 228.16447658771773 Training loss: 1.0844 Explore P:
0.0001
Episode: 385 Total reward: 270.6568153976807 Training loss: 3.7529 Explore P:
```

```
0.0001
Episode: 386 Total reward: 225.34379068169298 Training loss: 11.4236 Explore P:
0.0001
Episode: 387 Total reward: 6.650094454627208 Training loss: 1.0724 Explore P:
0.0001
Episode: 388 Total reward: 270.14909612957445 Training loss: 1.4295 Explore P:
0.0001
Episode: 389 Total reward: 212.12448445336156 Training loss: 1.1044 Explore P:
0.0001
Episode: 390 Total reward: 214.1443699988455 Training loss: 6.8448 Explore P:
0.0001
Episode: 391 Total reward: 249.75038809589842 Training loss: 7.0559 Explore P:
0.0001
Episode: 392 Total reward: 240.50429752760238 Training loss: 10.0601 Explore P:
0.0001
Episode: 393 Total reward: 249.88316237867005 Training loss: 5.4096 Explore P:
0.0001
Episode: 394 Total reward: 230.95539681806713 Training loss: 0.7386 Explore P:
0.0001
Episode: 395 Total reward: 239.59636344217992 Training loss: 1.1101 Explore P:
0.0001
Episode: 396 Total reward: 262.8887958386246 Training loss: 1.7269 Explore P:
0.0001
Episode: 397 Total reward: 273.01347207724865 Training loss: 5.3351 Explore P:
0.0001
Episode: 398 Total reward: 256.8935990143284 Training loss: 1.0753 Explore P:
0.0001
Episode: 399 Total reward: 259.38429968097756 Training loss: 1.3936 Explore P:
0.0001
```

Save policy network:

[35]:
```
torch.save(mainQN, log_path)
```

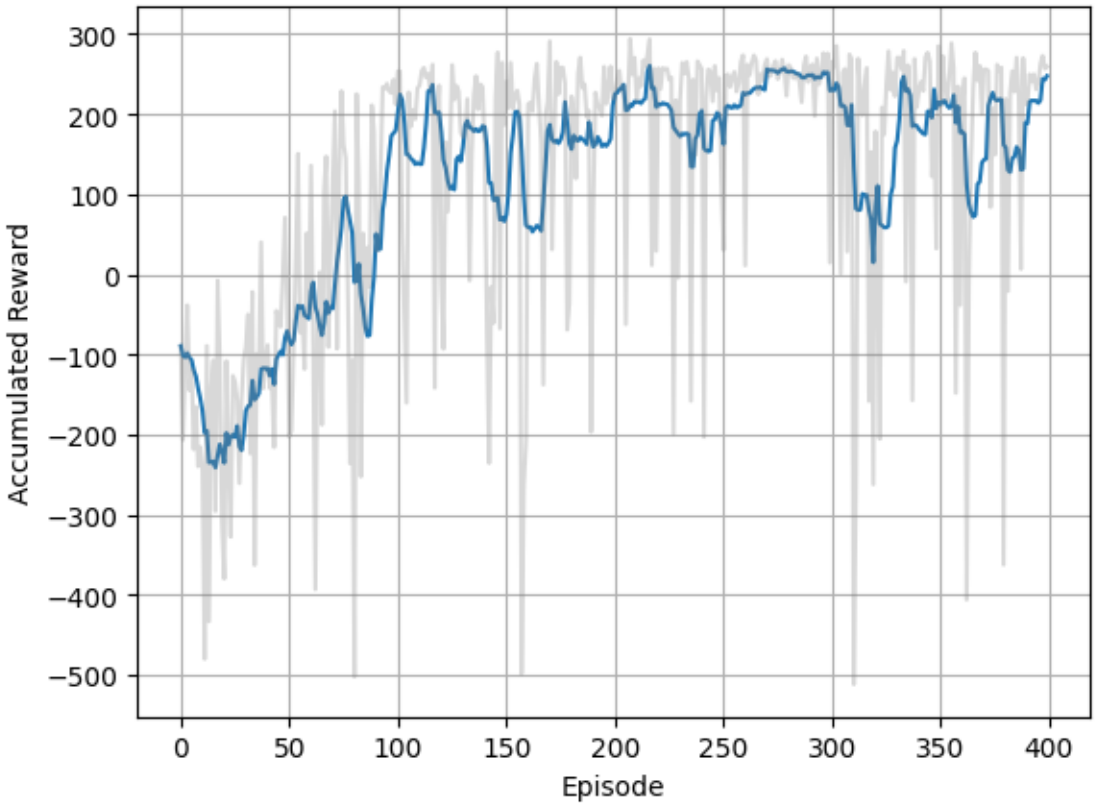
Plot learning process:

[36]:
```
# Moving average for smoothing plot
def running_mean(x, N):
    cumsum = np.cumsum(np.insert(x, 0, x[0]*np.ones(N)))
    return (cumsum[N:] - cumsum[:-N]) / N

eps, rews = np.array(total_reward_list).T
smoothed_rews = running_mean(rews, 10)

plt.plot(eps, smoothed_rews)
plt.grid()
plt.plot(eps, rews, color='grey', alpha=0.3)
plt.xlabel('Episode')
```

```
plt.ylabel('Accumulated Reward')
plt.savefig('deepQ.pdf')
```



Evaluate stored policy:

[ ]:
```
testQN = torch.load(log_path, weights_only=False)

test_episodes = 5

for ep in range(test_episodes):
    state = env_visual.reset()[0]
    print("initial state:", state)
    R = 0
    while True:
        # Get action from Q-network
        # Hm, the following line could perhaps be more elegant ...
        state_tensor = torch.from_numpy(np.resize(state, (1, state_size)).
 ↪astype(np.float32))
        Qs = testQN(state_tensor)
        action = torch.argmax(Qs).item()
```

```
        # Take action, get new state and reward
        next_state, reward, terminated, truncated, _ = env_visual.step(action)
        R += reward

        if terminated or truncated:
            print("reward:", R)
            break
        else:
            state = next_state
```

```
initial state: [ 0.00384388  1.4105263   0.38932493 -0.01751881 -0.00444726
-0.08818786
  0.          0.        ]
reward: 232.55242239220397
initial state: [-0.0020236   1.4059944  -0.20498946 -0.21892044  0.00235169
0.04643313
  0.          0.        ]
reward: 211.14939198092992
initial state: [ 0.00691957  1.4121294   0.70086545  0.05372036 -0.00801129
-0.15875655
  0.          0.        ]
reward: 210.58567907478857
initial state: [-0.00749578  1.411154   -0.7592486   0.01037014  0.00869245
0.17198119
  0.          0.        ]
reward: -375.47923794869416
initial state: [-0.00539989  1.4050679  -0.5469643  -0.26011088  0.0062639
0.1238956
  0.          0.        ]
reward: 209.28077355184206
```

```
The Kernel crashed while executing code in the current cell or a previous cell.

Please review the code in the cell(s) to identify a possible cause of the␣
↪failure.

Click <a href='https://aka.ms/vscodeJupyterKernelCrash'>here</a> for more info.

View Jupyter <a href='command:jupyter.viewOutput'>log</a> for further details.
```

```
[ ]:
```