James001

June 11, 2021

1 Crane V0 model training and simulations

Adapted from https://github.com/kinwo/deeprl-navigation (MIT License Copyright (c) 2018 Henry Chan)

Start Environement and create DQN Agent

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


     env = gym.make('crane-v0') #Load the environement
```

1.1 Agent

The DQN Agent adapted from https://github.com/kinwo/deeprl-navigation (MIT License Copyright (c) 2018 Henry Chan). The following block is also found in /model_script.py file in the current working directory

```
[2]: import numpy as np
     import random
     from collections import namedtuple, deque

     #from model_script import QNetwork # UNCOMMENT IF YOU ARE NOT IN A JUPYTER
      ↪NOTEBOOK

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

     BUFFER_SIZE = int(1e5)                  # replay buffer size
     BATCH_SIZE = 64                         # minibatch size
     GAMMA = 0.995           #was 0.99       # discount factor
     TAU = 1e-3                              # for soft update of target parameters
     LR = 5e-4                               # learning rate
     UPDATE_EVERY = 4                        # how often to update the network

     device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
     #device = torch.device('cuda:0')
```

```python
class Agent():
    """Interacts with and learns from the environment."""

    def __init__(self, state_size, action_size, seed):
        """Initialize an Agent object.

        Params
        ======
            state_size (int): dimension of each state
            action_size (int): dimension of each action
            seed (int): random seed
        """


        self.state_size = state_size
        self.action_size = action_size
        self.seed = random.seed(seed)

        # Q-Network
        self.qnetwork_local = QNetwork(state_size, action_size, seed).to(device)
        self.qnetwork_target = QNetwork(state_size, action_size, seed).
→to(device)
        self.optimizer = optim.Adam(self.qnetwork_local.parameters(), lr=LR)

        # Replay memory
        self.memory = ReplayBuffer(action_size, BUFFER_SIZE, BATCH_SIZE, seed)
        # Initialize time step (for updating every UPDATE_EVERY steps)
        self.t_step = 0

    def step(self, state, action, reward, next_state, done):
        # Save experience in replay memory
        self.memory.add(state, action, reward, next_state, done)

        # Learn every UPDATE_EVERY time steps.
        self.t_step = (self.t_step + 1) % UPDATE_EVERY
        if self.t_step == 0:
            # If enough samples are available in memory, get random subset and␣
→learn
            if len(self.memory) > BATCH_SIZE:
                experiences = self.memory.sample()
                self.learn(experiences, GAMMA)

    def act(self, state, eps=0.):
        """Returns actions for given state as per current policy.

        Params
```

```python
        ======
            state (array_like): current state
            eps (float): epsilon, for epsilon-greedy action selection
        """
        state = torch.from_numpy(state).float().unsqueeze(0).to(device)
        self.qnetwork_local.eval()
        with torch.no_grad():
            action_values = self.qnetwork_local(state)
        self.qnetwork_local.train()

        # Epsilon-greedy action selection
        if random.random() > eps:
            return np.argmax(action_values.cpu().data.numpy())
        else:
            return random.choice(np.arange(self.action_size))

    def learn(self, experiences, gamma):
        """Update value parameters using given batch of experience tuples.

        Params
        ======
            experiences (Tuple[torch.Variable]): tuple of (s, a, r, s', done)
→tuples
            gamma (float): discount factor
        """
        states, actions, rewards, next_states, dones = experiences

        # Get max predicted Q values (for next states) from target model
        Q_targets_next = self.qnetwork_target(next_states).detach().max(1)[0].
→unsqueeze(1)
        # Compute Q targets for current states
        Q_targets = rewards + (gamma * Q_targets_next * (1 - dones))

        # Get expected Q values from local model
        Q_expected = self.qnetwork_local(states).gather(1, actions)

        # Compute loss
        loss = F.mse_loss(Q_expected, Q_targets)
        # Minimize the loss
        self.optimizer.zero_grad()
        loss.backward()
        self.optimizer.step()

        # ------------------- update target network ------------------- #
        self.soft_update(self.qnetwork_local, self.qnetwork_target, TAU)

    def soft_update(self, local_model, target_model, tau):
```

```python
        """Soft update model parameters.
        _target =  * _local + (1 -  )* _target

        Params
        ======
            local_model (PyTorch model): weights will be copied from
            target_model (PyTorch model): weights will be copied to
            tau (float): interpolation parameter
        """
        for target_param, local_param in zip(target_model.parameters(),
→local_model.parameters()):
            target_param.data.copy_(tau*local_param.data + (1.
→0-tau)*target_param.data)


class ReplayBuffer:
    """Fixed-size buffer to store experience tuples."""

    def __init__(self, action_size, buffer_size, batch_size, seed):
        """Initialize a ReplayBuffer object.

        Params
        ======
            action_size (int): dimension of each action
            buffer_size (int): maximum size of buffer
            batch_size (int): size of each training batch
            seed (int): random seed
        """

        self.action_size = action_size
        self.memory = deque(maxlen=buffer_size)
        self.batch_size = batch_size
        self.experience = namedtuple("Experience", field_names=["state",
→"action", "reward", "next_state", "done"])
        self.seed = random.seed(seed)

    def add(self, state, action, reward, next_state, done):
        """Add a new experience to memory."""
        e = self.experience(state, action, reward, next_state, done)
        self.memory.append(e)

    def sample(self):
        """Randomly sample a batch of experiences from memory."""
        experiences = random.sample(self.memory, k=self.batch_size)

        states = torch.from_numpy(np.vstack([e.state for e in experiences if e
→is not None])).float().to(device)
```

```
        actions = torch.from_numpy(np.vstack([e.action for e in experiences if␣
↪e is not None])).long().to(device)
        rewards = torch.from_numpy(np.vstack([e.reward for e in experiences if␣
↪e is not None])).float().to(device)
        next_states = torch.from_numpy(np.vstack([e.next_state for e in␣
↪experiences if e is not None])).float().to(device)
        dones = torch.from_numpy(np.vstack([e.done for e in experiences if e is␣
↪not None]).astype(np.uint8)).float().to(device)

        return (states, actions, rewards, next_states, dones)

    def __len__(self):
        """Return the current size of internal memory."""
        return len(self.memory)
```

1.2   Deep Q_Network Model

A 2 linear hidden layer of 64 nodes each is created, with relu activation function.

```
[3]: import torch
import torch.nn as nn
import torch.nn.functional as F

class QNetwork(nn.Module):
    """Actor (Policy) Model."""

    def __init__(self, state_size, action_size, seed, fc1_units=64,␣
↪fc2_units=64):
        """Initialize parameters and build model.
        Params
        ======
            state_size (int): Dimension of each state
            action_size (int): Dimension of each action
            seed (int): Random seed
            fc1_units (int): Number of nodes in first hidden layer
            fc2_units (int): Number of nodes in second hidden layer
        """
        super(QNetwork, self).__init__()
        self.seed = torch.manual_seed(seed)
        self.fc1 = nn.Linear(state_size, fc1_units)
        self.fc2 = nn.Linear(fc1_units, fc2_units)
        self.fc3 = nn.Linear(fc2_units, action_size)

    def forward(self, state):
        """Build a network that maps state -> action values."""
        x = F.relu(self.fc1(state))
        x = F.relu(self.fc2(x))
```

```
        return self.fc3(x)
```

1.3 DQN Agent Training

1.3.1 Create DQN Agent

```
[4]: import torch
     import time
     from collections import deque

     #from agent_script import Agent   # UNCOMMENT IF YOU ARE NOT IN A JUPYTER␣
     ↪NOTEBOOK

     import matplotlib.pyplot as plt
     %matplotlib inline

     state_size=4
     action_size=3
     seed=0

     agent = Agent(state_size=4, action_size=3, seed=0)
```

In order to know when the environement is solved, we compute the moving score (the total rewards per episode) average over the last 100 episodes. If the moving average is over a chosen thershold (target_scores), the model is then saved to 'model_weight_name'.

For the Crane_v0 environement, the target score is 100 000, since it is the reward obtained by the agent when finding the flag.

```
[5]: model_weight_name = 'checkpoint_precise_2.pth'


     def dqn(n_episodes=10000, max_t=2000, eps_start=1.0, eps_end=0.01, eps_decay=0.
     ↪997, target_scores=100000.0):
         """Deep Q-Learning.

         Params
         ======
             n_episodes (int): maximum number of training episodes
             max_t (int): maximum number of timesteps per episode
             eps_start (float): starting value of epsilon, for epsilon-greedy action␣
     ↪selection
             eps_end (float): minimum value of epsilon
             eps_decay (float): multiplicative factor (per episode) for decreasing␣
     ↪epsilon
             target_scores (float): average scores aming to achieve, the agent will␣
     ↪stop training once it reaches this scores
```

```python
    """
    start = time.time()                  # Start time
    scores = []                          # list containing scores from each
↪episode
    scores_window = deque(maxlen=100)   # last 100 scores
    eps = eps_start                      # initialize epsilon

    for i_episode in range(1, n_episodes+1):
        # Reset env and score at the beginning of episode
        env_info = env.reset()                              # reset the
↪environment
        state = env.state                                   # get the current
↪state
        score = 0                                           # initialize the
↪score

        for t in range(max_t):
            action = agent.act(state, eps)
            env_info = env.step(action)                     # send the action to
↪the environment
            next_state = env_info[0]                        # get the next state
            reward = env_info[1]                            # get the reward
            done = env_info[2]                              # see if episode has
↪finished

            agent.step(state, action, reward, next_state, done)
            state = next_state
            score += reward
            if done:
                print("    Episode finished after {} timesteps".format(t+1))
                #print("final state is :", state)
                #print("Reward is : ", score)
                break

        scores_window.append(score)       # save most recent score
        scores.append(score)              # save most recent score
        eps = max(eps_end, eps_decay*eps) # decrease epsilon

        print('\rEpisode {}\tAverage Score: {:.2f}'.format(i_episode, np.
↪mean(scores_window)), end="")

        if i_episode % 100 == 0:
            print('\rEpisode {}\tAverage Score: {:.2f}'.format(i_episode, np.
↪mean(scores_window)))

        if np.mean(scores_window)>=target_scores:
```

```
            print('\nEnvironment solved in {:d} episodes!\tAverage Score: {:.
↪2f}'.format(i_episode, np.mean(scores_window)))
            torch.save(agent.qnetwork_local.state_dict(), model_weight_name)
            break

    time_elapsed = time.time() - start
    print("Time Elapse: {:.2f}".format(time_elapsed))

    return scores


scores = dqn(n_episodes=2000, max_t=1500, eps_start=1.0, eps_end=0.01,␣
↪eps_decay=0.997, target_scores=100000.0)
```

```
Episode 100     Average Score: -4280.83
Episode 200     Average Score: -1434.59
Episode 232     Average Score: 1968.738    Episode finished after 705 timesteps
Episode 264     Average Score: 5094.82     Episode finished after 623 timesteps
Episode 273     Average Score: 6476.31     Episode finished after 251 timesteps
Episode 300     Average Score: 7440.57
Episode 316     Average Score: 6790.40     Episode finished after 310 timesteps
Episode 320     Average Score: 7800.94     Episode finished after 1485 timesteps
Episode 321     Average Score: 8737.19     Episode finished after 589 timesteps
Episode 328     Average Score: 9867.86     Episode finished after 1010 timesteps
Episode 354     Average Score: 10126.66    Episode finished after 629 timesteps
Episode 364     Average Score: 12208.22    Episode finished after 681 timesteps
Episode 400     Average Score: 13053.45
Episode 415     Average Score: 13356.43    Episode finished after 1103 timesteps
Episode 418     Average Score: 13778.51    Episode finished after 839 timesteps
Episode 430     Average Score: 11929.81    Episode finished after 582 timesteps
Episode 431     Average Score: 13016.34    Episode finished after 1045 timesteps
Episode 443     Average Score: 14651.57    Episode finished after 1321 timesteps
Episode 444     Average Score: 15915.44    Episode finished after 982 timesteps
Episode 445     Average Score: 16988.54    Episode finished after 962 timesteps
Episode 448     Average Score: 18144.50    Episode finished after 488 timesteps
Episode 481     Average Score: 17383.89    Episode finished after 1102 timesteps
Episode 482     Average Score: 18254.33    Episode finished after 1353 timesteps
Episode 483     Average Score: 19354.75    Episode finished after 639 timesteps
Episode 484     Average Score: 20468.21    Episode finished after 1116 timesteps
Episode 495     Average Score: 22203.83    Episode finished after 296 timesteps
Episode 500     Average Score: 23639.19
Episode 544     Average Score: 22190.52    Episode finished after 1490 timesteps
Episode 600     Average Score: 18213.24
Episode 607     Average Score: 17693.67    Episode finished after 148 timesteps
Episode 633     Average Score: 16447.75    Episode finished after 1362 timesteps
Episode 663     Average Score: 14165.85    Episode finished after 1343 timesteps
Episode 686     Average Score: 14110.01    Episode finished after 1185 timesteps
Episode 700     Average Score: 16093.09
```

```
Episode 706     Average Score: 17342.40    Episode finished after 692 timesteps
Episode 728     Average Score: 21077.50    Episode finished after 454 timesteps
Episode 732     Average Score: 22622.20    Episode finished after 569 timesteps
Episode 733     Average Score: 23516.84    Episode finished after 751 timesteps
Episode 746     Average Score: 25769.87    Episode finished after 1322 timesteps
Episode 747     Average Score: 27087.13    Episode finished after 1095 timesteps
Episode 751     Average Score: 28839.76    Episode finished after 1033 timesteps
Episode 762     Average Score: 31443.02    Episode finished after 1495 timesteps
Episode 764     Average Score: 31576.69    Episode finished after 1459 timesteps
Episode 767     Average Score: 33245.21    Episode finished after 1065 timesteps
Episode 776     Average Score: 35590.93    Episode finished after 1377 timesteps
Episode 777     Average Score: 36909.77    Episode finished after 983 timesteps
Episode 780     Average Score: 37774.89    Episode finished after 666 timesteps
Episode 781     Average Score: 38646.14    Episode finished after 619 timesteps
Episode 782     Average Score: 39508.82    Episode finished after 403 timesteps
Episode 786     Average Score: 40583.06    Episode finished after 333 timesteps
Episode 787     Average Score: 40534.85    Episode finished after 1449 timesteps
Episode 788     Average Score: 41920.65    Episode finished after 410 timesteps
Episode 789     Average Score: 42712.79    Episode finished after 479 timesteps
Episode 790     Average Score: 43460.00    Episode finished after 402 timesteps
Episode 791     Average Score: 44226.70    Episode finished after 382 timesteps
Episode 792     Average Score: 44990.60    Episode finished after 666 timesteps
Episode 793     Average Score: 45827.58    Episode finished after 692 timesteps
Episode 795     Average Score: 46701.95    Episode finished after 1267 timesteps
Episode 796     Average Score: 47941.82    Episode finished after 1416 timesteps
Episode 800     Average Score: 49045.14
Episode 804     Average Score: 48987.04    Episode finished after 713 timesteps
Episode 815     Average Score: 48805.46    Episode finished after 790 timesteps
Episode 816     Average Score: 49694.09    Episode finished after 442 timesteps
Episode 821     Average Score: 50686.48    Episode finished after 455 timesteps
Episode 822     Average Score: 51501.36    Episode finished after 532 timesteps
Episode 823     Average Score: 52326.65    Episode finished after 365 timesteps
Episode 824     Average Score: 53104.06    Episode finished after 795 timesteps
Episode 825     Average Score: 54043.83    Episode finished after 1059 timesteps
Episode 829     Average Score: 54243.01    Episode finished after 1187 timesteps
Episode 830     Average Score: 55207.94    Episode finished after 751 timesteps
Episode 831     Average Score: 56019.97    Episode finished after 782 timesteps
Episode 832     Average Score: 56883.01    Episode finished after 1300 timesteps
Episode 834     Average Score: 56055.01    Episode finished after 1394 timesteps
Episode 835     Average Score: 57050.12    Episode finished after 464 timesteps
Episode 836     Average Score: 57820.68    Episode finished after 511 timesteps
Episode 842     Average Score: 58879.70    Episode finished after 813 timesteps
Episode 843     Average Score: 59760.17    Episode finished after 295 timesteps
Episode 844     Average Score: 60556.57    Episode finished after 347 timesteps
Episode 850     Average Score: 59481.41    Episode finished after 600 timesteps
Episode 852     Average Score: 59310.15    Episode finished after 1432 timesteps
Episode 865     Average Score: 58624.85    Episode finished after 724 timesteps
Episode 867     Average Score: 59435.09    Episode finished after 934 timesteps
```

```
Episode 868      Average Score: 59467.61    Episode finished after 1450 timesteps
Episode 871      Average Score: 60652.23    Episode finished after 351 timesteps
Episode 873      Average Score: 61442.30    Episode finished after 306 timesteps
Episode 875      Average Score: 62296.39    Episode finished after 311 timesteps
Episode 877      Average Score: 62118.29    Episode finished after 1127 timesteps
Episode 878      Average Score: 62213.97    Episode finished after 1239 timesteps
Episode 879      Average Score: 63213.89    Episode finished after 489 timesteps
Episode 880      Average Score: 64053.02    Episode finished after 305 timesteps
Episode 881      Average Score: 63975.99    Episode finished after 1450 timesteps
Episode 885      Average Score: 63527.67    Episode finished after 282 timesteps
Episode 886      Average Score: 64216.61    Episode finished after 334 timesteps
Episode 888      Average Score: 63206.87    Episode finished after 355 timesteps
Episode 889      Average Score: 63195.09    Episode finished after 320 timesteps
Episode 890      Average Score: 63167.69    Episode finished after 855 timesteps
Episode 891      Average Score: 63312.24    Episode finished after 289 timesteps
Episode 893      Average Score: 62428.80    Episode finished after 718 timesteps
Episode 894      Average Score: 62441.16    Episode finished after 269 timesteps
Episode 895      Average Score: 63177.53    Episode finished after 162 timesteps
Episode 898      Average Score: 62039.15    Episode finished after 941 timesteps
Episode 899      Average Score: 62948.79    Episode finished after 864 timesteps
Episode 900      Average Score: 63870.74
Episode 901      Average Score: 63987.93    Episode finished after 1132 timesteps
Episode 902      Average Score: 64967.64    Episode finished after 298 timesteps
Episode 903      Average Score: 65760.24    Episode finished after 273 timesteps
Episode 904      Average Score: 66528.73    Episode finished after 1000 timesteps
Episode 905      Average Score: 66643.87    Episode finished after 297 timesteps
Episode 906      Average Score: 67494.66    Episode finished after 301 timesteps
Episode 907      Average Score: 68319.60    Episode finished after 183 timesteps
Episode 910      Average Score: 69191.43    Episode finished after 332 timesteps
Episode 911      Average Score: 69914.59    Episode finished after 907 timesteps
Episode 912      Average Score: 70798.28    Episode finished after 268 timesteps
Episode 918      Average Score: 69845.01    Episode finished after 301 timesteps
Episode 922      Average Score: 69928.87    Episode finished after 966 timesteps
Episode 923      Average Score: 70045.78    Episode finished after 1376 timesteps
Episode 924      Average Score: 70287.77    Episode finished after 325 timesteps
Episode 925      Average Score: 70166.61    Episode finished after 957 timesteps
Episode 927      Average Score: 70153.40    Episode finished after 163 timesteps
Episode 928      Average Score: 70885.51    Episode finished after 178 timesteps
Episode 929      Average Score: 71610.61    Episode finished after 150 timesteps
Episode 930      Average Score: 71382.08    Episode finished after 889 timesteps
Episode 931      Average Score: 71445.53    Episode finished after 818 timesteps
Episode 933      Average Score: 70498.86    Episode finished after 301 timesteps
Episode 935      Average Score: 70452.87    Episode finished after 187 timesteps
Episode 936      Average Score: 70394.06    Episode finished after 246 timesteps
Episode 937      Average Score: 70350.70    Episode finished after 209 timesteps
Episode 939      Average Score: 71016.48    Episode finished after 771 timesteps
Episode 942      Average Score: 71827.99    Episode finished after 230 timesteps
Episode 943      Average Score: 71713.35    Episode finished after 791 timesteps
```

```
Episode 947      Average Score: 71047.32    Episode finished after 284 timesteps
Episode 948      Average Score: 71789.64    Episode finished after 153 timesteps
Episode 949      Average Score: 72532.59    Episode finished after 338 timesteps
Episode 955      Average Score: 71636.72    Episode finished after 162 timesteps
Episode 956      Average Score: 72352.18    Episode finished after 1152 timesteps
Episode 957      Average Score: 73280.53    Episode finished after 686 timesteps
Episode 959      Average Score: 74052.52    Episode finished after 367 timesteps
Episode 962      Average Score: 74641.20    Episode finished after 1127 timesteps
Episode 967      Average Score: 74696.30    Episode finished after 1074 timesteps
Episode 969      Average Score: 73676.95    Episode finished after 1026 timesteps
Episode 970      Average Score: 74511.22    Episode finished after 186 timesteps
Episode 972      Average Score: 74455.99    Episode finished after 576 timesteps
Episode 973      Average Score: 75216.22    Episode finished after 151 timesteps
Episode 976      Average Score: 74426.71    Episode finished after 433 timesteps
Episode 977      Average Score: 75163.19    Episode finished after 1066 timesteps
Episode 980      Average Score: 73392.64    Episode finished after 396 timesteps
Episode 981      Average Score: 73419.19    Episode finished after 308 timesteps
Episode 983      Average Score: 73052.34    Episode finished after 152 timesteps
Episode 989      Average Score: 71351.60    Episode finished after 729 timesteps
Episode 990      Average Score: 71457.47    Episode finished after 682 timesteps
Episode 991      Average Score: 71426.86    Episode finished after 738 timesteps
Episode 994      Average Score: 70728.21    Episode finished after 603 timesteps
Episode 996      Average Score: 70084.07    Episode finished after 133 timesteps
Episode 999      Average Score: 69878.16    Episode finished after 156 timesteps
Episode 1000     Average Score: 69715.03
    Episode finished after 273 timesteps
Episode 1001     Average Score: 70415.56    Episode finished after 357 timesteps
Episode 1002     Average Score: 70217.90    Episode finished after 320 timesteps
Episode 1003     Average Score: 70219.49    Episode finished after 198 timesteps
Episode 1004     Average Score: 70197.63    Episode finished after 434 timesteps
Episode 1005     Average Score: 70021.01    Episode finished after 896 timesteps
Episode 1006     Average Score: 70168.65    Episode finished after 1495 timesteps
Episode 1007     Average Score: 70455.15    Episode finished after 1206 timesteps
Episode 1009     Average Score: 70709.16    Episode finished after 627 timesteps
Episode 1010     Average Score: 71497.28    Episode finished after 703 timesteps
Episode 1011     Average Score: 71586.54    Episode finished after 670 timesteps
Episode 1013     Average Score: 70829.94    Episode finished after 827 timesteps
Episode 1014     Average Score: 71674.31    Episode finished after 947 timesteps
Episode 1015     Average Score: 72681.81    Episode finished after 383 timesteps
Episode 1020     Average Score: 72523.28    Episode finished after 990 timesteps
Episode 1021     Average Score: 73385.80    Episode finished after 349 timesteps
Episode 1022     Average Score: 74061.19    Episode finished after 429 timesteps
Episode 1024     Average Score: 72868.30    Episode finished after 144 timesteps
Episode 1025     Average Score: 72830.33    Episode finished after 1293 timesteps
Episode 1026     Average Score: 72835.66    Episode finished after 400 timesteps
Episode 1031     Average Score: 70517.93    Episode finished after 733 timesteps
Episode 1034     Average Score: 69722.84    Episode finished after 376 timesteps
Episode 1035     Average Score: 70457.88    Episode finished after 391 timesteps
```

```
Episode 1036    Average Score: 70489.37    Episode finished after 1476 timesteps
Episode 1037    Average Score: 70689.51    Episode finished after 644 timesteps
Episode 1041    Average Score: 69990.85    Episode finished after 1031 timesteps
Episode 1044    Average Score: 69179.31    Episode finished after 229 timesteps
Episode 1046    Average Score: 69877.82    Episode finished after 499 timesteps
Episode 1048    Average Score: 69924.13    Episode finished after 888 timesteps
Episode 1049    Average Score: 70081.13    Episode finished after 1176 timesteps
Episode 1051    Average Score: 70259.90    Episode finished after 798 timesteps
Episode 1052    Average Score: 71051.22    Episode finished after 671 timesteps
Episode 1053    Average Score: 71885.81    Episode finished after 546 timesteps
Episode 1054    Average Score: 72672.21    Episode finished after 652 timesteps
Episode 1057    Average Score: 71826.99    Episode finished after 306 timesteps
Episode 1058    Average Score: 71722.60    Episode finished after 930 timesteps
Episode 1062    Average Score: 72095.70    Episode finished after 1495 timesteps
Episode 1063    Average Score: 72150.01    Episode finished after 496 timesteps
Episode 1065    Average Score: 72814.10    Episode finished after 663 timesteps
Episode 1066    Average Score: 73636.35    Episode finished after 1027 timesteps
Episode 1068    Average Score: 73633.65    Episode finished after 444 timesteps
Episode 1069    Average Score: 74446.40    Episode finished after 1250 timesteps
Episode 1071    Average Score: 73794.05    Episode finished after 468 timesteps
Episode 1072    Average Score: 74507.64    Episode finished after 370 timesteps
Episode 1075    Average Score: 73730.46    Episode finished after 1234 timesteps
Episode 1076    Average Score: 74694.64    Episode finished after 463 timesteps
Episode 1077    Average Score: 74687.06    Episode finished after 682 timesteps
Episode 1078    Average Score: 74583.85    Episode finished after 550 timesteps
Episode 1079    Average Score: 75361.56    Episode finished after 475 timesteps
Episode 1080    Average Score: 76065.17    Episode finished after 548 timesteps
Episode 1084    Average Score: 74497.57    Episode finished after 630 timesteps
Episode 1085    Average Score: 75275.81    Episode finished after 210 timesteps
Episode 1086    Average Score: 76103.72    Episode finished after 481 timesteps
Episode 1087    Average Score: 76861.61    Episode finished after 431 timesteps
Episode 1088    Average Score: 77557.62    Episode finished after 225 timesteps
Episode 1090    Average Score: 77482.54    Episode finished after 902 timesteps
Episode 1091    Average Score: 77510.43    Episode finished after 838 timesteps
Episode 1094    Average Score: 77438.46    Episode finished after 587 timesteps
Episode 1096    Average Score: 77426.44    Episode finished after 367 timesteps
Episode 1097    Average Score: 77456.22    Episode finished after 445 timesteps
Episode 1100    Average Score: 77385.85
Episode 1101    Average Score: 76644.07    Episode finished after 304 timesteps
Episode 1102    Average Score: 76626.00    Episode finished after 567 timesteps
Episode 1106    Average Score: 74315.58    Episode finished after 724 timesteps
Episode 1107    Average Score: 74067.72    Episode finished after 494 timesteps
Episode 1108    Average Score: 73843.04    Episode finished after 314 timesteps
Episode 1109    Average Score: 74564.58    Episode finished after 350 timesteps
Episode 1111    Average Score: 73656.35    Episode finished after 873 timesteps
Episode 1112    Average Score: 73646.85    Episode finished after 727 timesteps
Episode 1113    Average Score: 74386.98    Episode finished after 634 timesteps
Episode 1114    Average Score: 74342.47    Episode finished after 633 timesteps
```

```
Episode 1115    Average Score: 74208.32    Episode finished after 357 timesteps
Episode 1116    Average Score: 74195.57    Episode finished after 515 timesteps
Episode 1117    Average Score: 74929.94    Episode finished after 371 timesteps
Episode 1121    Average Score: 74444.03    Episode finished after 1231 timesteps
Episode 1122    Average Score: 74614.25    Episode finished after 969 timesteps
Episode 1123    Average Score: 74725.96    Episode finished after 1338 timesteps
Episode 1124    Average Score: 75678.68    Episode finished after 722 timesteps
Episode 1125    Average Score: 75730.14    Episode finished after 493 timesteps
Episode 1126    Average Score: 75547.12    Episode finished after 1138 timesteps
Episode 1127    Average Score: 75677.87    Episode finished after 580 timesteps
Episode 1128    Average Score: 76418.58    Episode finished after 616 timesteps
Episode 1129    Average Score: 77175.69    Episode finished after 858 timesteps
Episode 1130    Average Score: 77992.80    Episode finished after 251 timesteps
Episode 1131    Average Score: 78702.77    Episode finished after 632 timesteps
Episode 1133    Average Score: 78612.73    Episode finished after 505 timesteps
Episode 1135    Average Score: 78297.40    Episode finished after 781 timesteps
Episode 1136    Average Score: 78370.31    Episode finished after 214 timesteps
Episode 1137    Average Score: 78143.83    Episode finished after 667 timesteps
Episode 1138    Average Score: 78081.25    Episode finished after 647 timesteps
Episode 1141    Average Score: 78515.19    Episode finished after 634 timesteps
Episode 1142    Average Score: 78453.41    Episode finished after 578 timesteps
Episode 1143    Average Score: 79259.37    Episode finished after 570 timesteps
Episode 1144    Average Score: 80085.52    Episode finished after 1036 timesteps
Episode 1145    Average Score: 80225.61    Episode finished after 475 timesteps
Episode 1146    Average Score: 80978.95    Episode finished after 663 timesteps
Episode 1148    Average Score: 80974.81    Episode finished after 598 timesteps
Episode 1149    Average Score: 80856.46    Episode finished after 553 timesteps
Episode 1150    Average Score: 80641.38    Episode finished after 603 timesteps
Episode 1152    Average Score: 80312.20    Episode finished after 1371 timesteps
Episode 1154    Average Score: 79308.01    Episode finished after 616 timesteps
Episode 1155    Average Score: 79240.29    Episode finished after 1381 timesteps
Episode 1157    Average Score: 79981.46    Episode finished after 1331 timesteps
Episode 1158    Average Score: 79963.16    Episode finished after 1081 timesteps
Episode 1159    Average Score: 79989.08    Episode finished after 1087 timesteps
Episode 1160    Average Score: 80872.33    Episode finished after 1465 timesteps
Episode 1162    Average Score: 81560.93    Episode finished after 583 timesteps
Episode 1163    Average Score: 81356.87    Episode finished after 955 timesteps
Episode 1165    Average Score: 81227.58    Episode finished after 737 timesteps
Episode 1166    Average Score: 81147.49    Episode finished after 1302 timesteps
Episode 1167    Average Score: 81192.42    Episode finished after 702 timesteps
Episode 1169    Average Score: 81263.16    Episode finished after 690 timesteps
Episode 1170    Average Score: 81087.92    Episode finished after 400 timesteps
Episode 1171    Average Score: 81864.63    Episode finished after 392 timesteps
Episode 1172    Average Score: 81873.24    Episode finished after 969 timesteps
Episode 1173    Average Score: 81992.66    Episode finished after 397 timesteps
Episode 1174    Average Score: 82725.90    Episode finished after 1214 timesteps
Episode 1176    Average Score: 82493.42    Episode finished after 578 timesteps
Episode 1177    Average Score: 82520.11    Episode finished after 715 timesteps
```
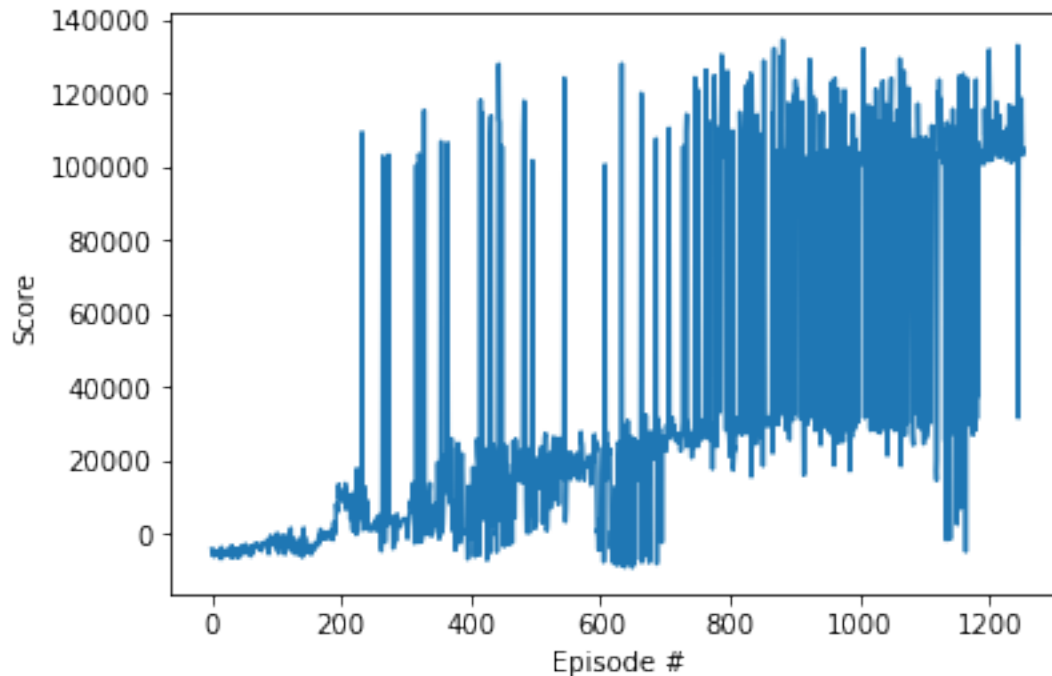
```
Episode 1178     Average Score: 82532.65     Episode finished after 362 timesteps
Episode 1180     Average Score: 81747.64     Episode finished after 1206 timesteps
Episode 1182     Average Score: 81983.20     Episode finished after 531 timesteps
Episode 1184     Average Score: 82915.33     Episode finished after 476 timesteps
Episode 1185     Average Score: 82904.81     Episode finished after 407 timesteps
Episode 1186     Average Score: 82945.44     Episode finished after 315 timesteps
Episode 1187     Average Score: 82920.86     Episode finished after 355 timesteps
Episode 1188     Average Score: 82920.94     Episode finished after 264 timesteps
Episode 1189     Average Score: 82933.35     Episode finished after 337 timesteps
Episode 1190     Average Score: 83649.71     Episode finished after 270 timesteps
Episode 1191     Average Score: 83497.55     Episode finished after 251 timesteps
Episode 1192     Average Score: 83451.63     Episode finished after 221 timesteps
Episode 1193     Average Score: 84167.17     Episode finished after 825 timesteps
Episode 1194     Average Score: 85033.46     Episode finished after 600 timesteps
Episode 1195     Average Score: 85041.36     Episode finished after 896 timesteps
Episode 1196     Average Score: 85898.67     Episode finished after 595 timesteps
Episode 1197     Average Score: 85951.55     Episode finished after 542 timesteps
Episode 1198     Average Score: 86013.89     Episode finished after 446 timesteps
Episode 1199     Average Score: 86776.51     Episode finished after 342 timesteps
Episode 1200     Average Score: 87546.29
     Episode finished after 1317 timesteps
Episode 1201     Average Score: 88578.68     Episode finished after 1084 timesteps
Episode 1202     Average Score: 88772.01     Episode finished after 316 timesteps
Episode 1203     Average Score: 88708.99     Episode finished after 531 timesteps
Episode 1204     Average Score: 89490.03     Episode finished after 590 timesteps
Episode 1205     Average Score: 90234.84     Episode finished after 455 timesteps
Episode 1206     Average Score: 90942.48     Episode finished after 273 timesteps
Episode 1207     Average Score: 90892.78     Episode finished after 667 timesteps
Episode 1208     Average Score: 90978.47     Episode finished after 588 timesteps
Episode 1209     Average Score: 91019.12     Episode finished after 393 timesteps
Episode 1210     Average Score: 91034.01     Episode finished after 799 timesteps
Episode 1211     Average Score: 91848.18     Episode finished after 943 timesteps
Episode 1212     Average Score: 91913.52     Episode finished after 838 timesteps
Episode 1213     Average Score: 91967.11     Episode finished after 567 timesteps
Episode 1214     Average Score: 91913.78     Episode finished after 483 timesteps
Episode 1215     Average Score: 91920.58     Episode finished after 979 timesteps
Episode 1216     Average Score: 91935.25     Episode finished after 357 timesteps
Episode 1217     Average Score: 91935.67     Episode finished after 619 timesteps
Episode 1218     Average Score: 92039.69     Episode finished after 301 timesteps
Episode 1219     Average Score: 92893.80     Episode finished after 457 timesteps
Episode 1220     Average Score: 93777.94     Episode finished after 409 timesteps
Episode 1221     Average Score: 94615.85     Episode finished after 524 timesteps
Episode 1222     Average Score: 94470.94     Episode finished after 764 timesteps
Episode 1223     Average Score: 94423.87     Episode finished after 679 timesteps
Episode 1224     Average Score: 94270.32     Episode finished after 470 timesteps
Episode 1225     Average Score: 94271.55     Episode finished after 629 timesteps
Episode 1226     Average Score: 94335.08     Episode finished after 310 timesteps
Episode 1227     Average Score: 94168.58     Episode finished after 354 timesteps
```

```
Episode 1228      Average Score: 94155.41      Episode finished after 438 timesteps
Episode 1229      Average Score: 94169.19      Episode finished after 729 timesteps
Episode 1230      Average Score: 94178.12      Episode finished after 577 timesteps
Episode 1231      Average Score: 94226.62      Episode finished after 860 timesteps
Episode 1232      Average Score: 94338.42      Episode finished after 506 timesteps
Episode 1233      Average Score: 95128.12      Episode finished after 498 timesteps
Episode 1234      Average Score: 95138.56      Episode finished after 350 timesteps
Episode 1235      Average Score: 96165.91      Episode finished after 381 timesteps
Episode 1236      Average Score: 96099.95      Episode finished after 586 timesteps
Episode 1237      Average Score: 96155.49      Episode finished after 742 timesteps
Episode 1238      Average Score: 96274.61      Episode finished after 441 timesteps
Episode 1239      Average Score: 96285.93      Episode finished after 417 timesteps
Episode 1240      Average Score: 97016.48      Episode finished after 619 timesteps
Episode 1241      Average Score: 98048.04      Episode finished after 401 timesteps
Episode 1242      Average Score: 97973.76      Episode finished after 480 timesteps
Episode 1243      Average Score: 97911.61      Episode finished after 394 timesteps
Episode 1244      Average Score: 97884.83      Episode finished after 1363 timesteps
Episode 1246      Average Score: 97327.80      Episode finished after 311 timesteps
Episode 1247      Average Score: 97243.37      Episode finished after 423 timesteps
Episode 1248      Average Score: 98015.88      Episode finished after 582 timesteps
Episode 1249      Average Score: 98042.92      Episode finished after 445 timesteps
Episode 1250      Average Score: 98049.97      Episode finished after 968 timesteps
Episode 1251      Average Score: 98185.73      Episode finished after 600 timesteps
Episode 1252      Average Score: 99208.51      Episode finished after 415 timesteps
Episode 1253      Average Score: 99220.89      Episode finished after 664 timesteps
Episode 1254      Average Score: 100087.02
Environment solved in 1254 episodes!     Average Score: 100087.02
Time Elapse: 1705.75
```

1.3.2 Score plot for each episodes during training

```python
# plot the scores
fig = plt.figure()
ax = fig.add_subplot(111)
plt.plot(np.arange(len(scores)), scores)
plt.ylabel('Score')
plt.xlabel('Episode #')
plt.savefig('plots/model_training.png', dpi = 200) #UNCOMMENT TO SAVE PLOT
plt.show()
```

1.4 Analysis and inspections

1.4.1 Watch the agent running (Using saved weights)

```
[7]: # load the weights from file
     agent.qnetwork_local.load_state_dict(torch.load('checkpoint_precise_2.pth'))
     #agent.qnetwork_local.load_state_dict(torch.load('checkpoint_precise.pth'))


     env_info = env.reset()                              # reset the environment
     state = env.state                                   # get the current state
     score = 0                                           # initialize the score

     t = 0
     while True:
         env.render()
         time.sleep(0.008)
         action = agent.act(state)                       # select an action
         env_info = env.step(action)                     # send the action to the␣
     ↪environment
         next_state = env_info[0]                         # get the next state
         reward = env_info[1]                             # get the reward
         done = env_info[2]                               # see if episode has finished
         score += reward                                 # update the score
```

16

```
    state = next_state  # roll over the state to next time step
    t += 1
    if done:                                        # exit loop if episode␣
 ↪finished
        print("final state is :", state)
        print("Total steps : ", t)
        break

print("Score: {}".format(score))
env.close()
```

```
final state is : [ 1.01893092e+00  2.26878363e-03  3.12455368e+00
-2.64702502e-02]
Total steps :   512
Score: 103213.0819224108
```

1.4.2 Calculate time steps and time before solving environement

```
[14]: #Choose model
      agent.qnetwork_local.load_state_dict(torch.load('checkpoint_precise.pth'))

      env_info = env.reset()                              # reset the environment
      state = env.state                                   # get the current state
      score = 0                                           # initialize the score

      t = 0
      while True:
          env.render()
          #time.sleep(0.02)                               # Actual time step
          action = agent.act(state)                       # select an action
          env_info = env.step(action)                     # send the action to the␣
      ↪environment
          next_state = env_info[0]                        # get the next state
          reward = env_info[1]                            # get the reward
          done = env_info[2]                              # see if episode has finished
          score += reward                                 # update the score
          state = next_state  # roll over the state to next time step
          t += 1
          if done:                                        # exit loop if episode␣
      ↪finished
              print("\r final state is :", state)
              print("\r Total steps : ", t)
              print("\r Total time is : ", env.tau * t)
              break

      print("Score: {}".format(score))
      env.close()
```

17

```
final state is : [ 0.99739885  0.0387832   3.17148444 -0.00326196]
Total steps :   393
Total time is :   7.86
Score: 105153.49475149131
```

1.4.3 Plot state graphs against time

```python
[15]: #Choose model
      agent.qnetwork_local.load_state_dict(torch.load('checkpoint_precise.pth'))
      agent.qnetwork_local.state_dict() #UNCOMMENT TO USE LAST TRAINED WEIGHTS


      env_info = env.reset()                           # reset the environment
      state = env.state                                # get the current state
      score = 0                                        # initialize the score

      arr_x = []
      arr_x_dot = []
      arr_theta = []
      arr_theta_dot = []
      arr_t = []


      t = 0.0
      while True:
          env.render()
          #time.sleep(0.008)
          action = agent.act(state)                    # select an action
          env_info = env.step(action)                  # send the action to the
       ↪environment
          next_state = env_info[0]                      # get the next state
          reward = env_info[1]                          # get the reward
          done = env_info[2]                            # see if episode has finished
          score += reward                               # update the score
          state = next_state  # roll over the state to next time step
          t += 1.0
          arr_t.append(t)
          arr_x.append(state[0])
          arr_x_dot.append(state[1])
          arr_theta.append(state[2])
          arr_theta_dot.append(state[3])

          if done:                                      # exit loop if episode
       ↪finished
              print("Total steps : ", t)
              print("Total time is : ", env.tau * t)
              break
      arr_t = 0.02*np.array(arr_t)
```

18

```
print("Score: {}".format(score))
env.close()
```

```
Total steps :   383.0
Total time is :   7.66
Score: 104468.93876048437
```

```
[16]: fig = plt.figure()
      ax = fig.add_subplot(111)
      plt.plot(arr_t, arr_x, label='Cart Position')
      plt.ylabel('X Position [m]')
      plt.xlabel('Time [s]')
      plt.title('Position / Time')
      plt.hlines(1.0, 0, arr_t[-1], colors='r', linestyles='solid', label='Goal␣
       ↪Position')
      plt.legend()
      plt.grid()
      #plt.savefig('plots/model_1_x.png', dpi = 200)
      plt.show()

      fig = plt.figure()
      ax = fig.add_subplot(111)
      plt.plot(arr_t, arr_x_dot, label = 'Cart Velocity')
      plt.ylabel('X Velocity [m / s]')
      plt.xlabel('Time [s]')
      plt.title('Velocity / Time')
      plt.hlines(0.0, 0, arr_t[-1], colors='r', linestyles='solid', label='Goal␣
       ↪Velocity')
      plt.legend()
      plt.grid()
      #plt.savefig('plots/model_1_x_dot.png', dpi = 200)
      plt.show()

      fig = plt.figure()
      ax = fig.add_subplot(111)
      plt.plot(arr_t, arr_theta, label = 'Pole angle')
      plt.ylabel('Theta [rad]')
      plt.xlabel('Time [s]')
      plt.title('Theta / Time')
      plt.hlines(np.pi, 0, arr_t[-1], colors='r', linestyles='solid', label='Goal␣
       ↪Angle')
      plt.legend()
      plt.grid()
      #plt.savefig('plots/model_1_theta.png', dpi = 200)
      plt.show()

      fig = plt.figure()
```
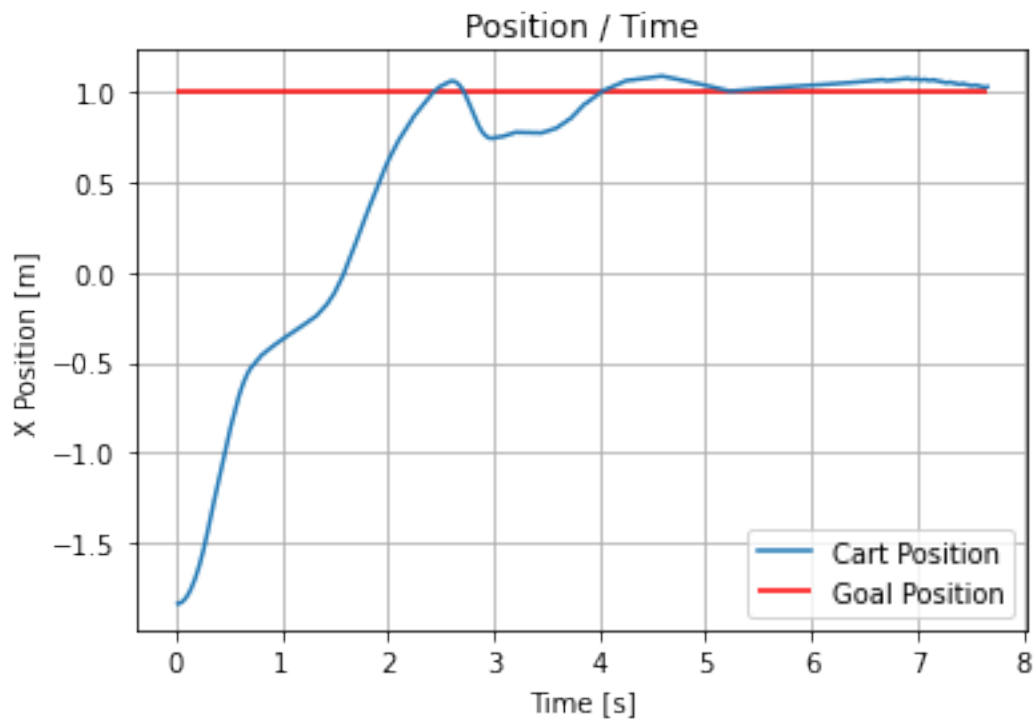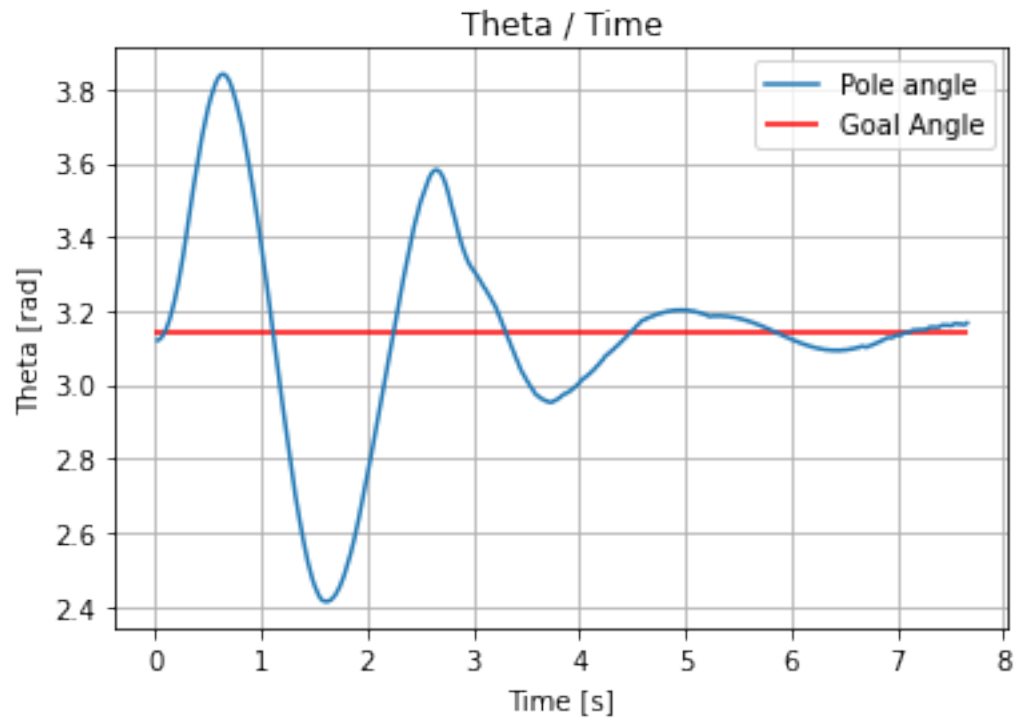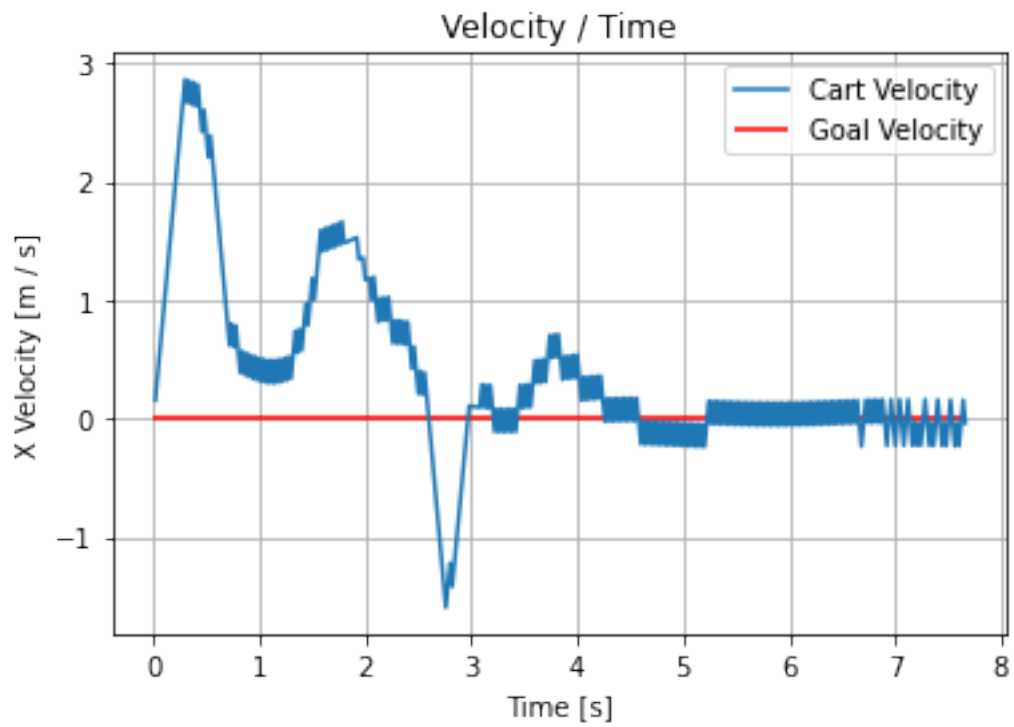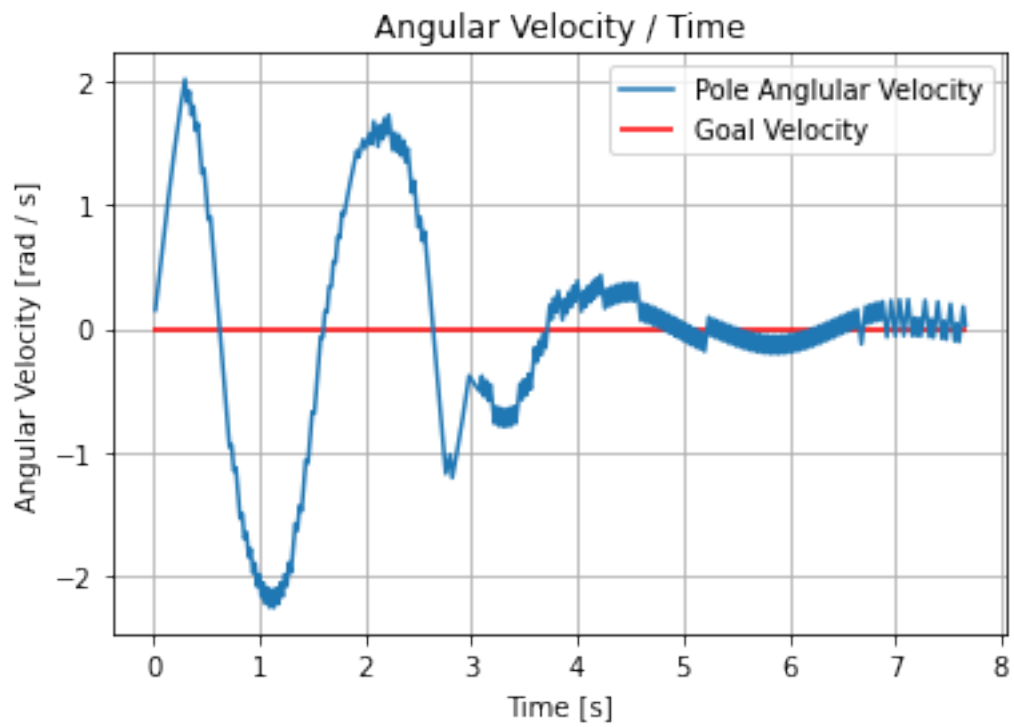
```
ax = fig.add_subplot(111)
plt.plot(arr_t, arr_theta_dot, label = 'Pole Anglular Velocity')
plt.ylabel('Angular Velocity [rad / s]')
plt.xlabel('Time [s]')
plt.title('Angular Velocity / Time')
plt.hlines(0.0, 0, arr_t[-1], colors='r', linestyles='solid', label='Goal␣
 ↪Velocity')
plt.legend()
plt.grid()
#plt.savefig('plots/model_1_theta_dot.png', dpi = 200)
plt.show()

initial_state = [arr_x[0], arr_x_dot[0], arr_theta[0], arr_theta_dot[0]]
print('initial state is : ', initial_state)
```

Velocity / Time

Theta / Time

Angular Velocity / Time

initial state is : [-1.8381194721045466, 0.1541213842217936, 3.12001952945492,
0.14986158839055636]