Self-Driving Cars

Ex. 01 - Coding Challenge - Imitation Learning

Katrin Renz
Autonomous Vision Group
University of Tübingen

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

MAX-PLANCK-GESELLSCHAFT

ellis
European Laboratory for Learning and Intelligent Systems

Questions?

Exercise Setup

Files

- `01_imitation_learning_exercise.pdf`
- `main.py`
- `network.py`, `training.py`, `demonstrations.py`
- `submission.txt`
- `data`, a folder with demonstrations

Exercise Setup

Files

- `01_imitation_learning_exercise.pdf`
- `main.py`
- `network.py`, `training.py`, `demonstrations.py`
- `submission.txt`
- `data`, a folder with demonstrations

Submit

- submission information: please fill out `submission.txt`
- your code: `network.py`, `training.py`, `demonstrations.py` as a `.zip` file
- your pre-trained model as a `.t7` file

Deadline: **Tue, 16. November 2021 - 8pm**

Exercise Setup

Do’s

- comment your code
- use docstrings
- use self-explanatory variable names
- structure your code well

Exercise Setup

Do's

- comment your code
- use docstrings
- use self-explanatory variable names
- structure your code well

Do not's

- change `main.py`, especially `calculate_score_for_leaderboard()`
- install more packages
- change the gym environment

Imitation Learning

Behavioral Cloning

Imitation Learning

Components:

- State: $s \in \mathcal{S}$ — may be partially observed (e.g., game screen)
- Action: $a \in \mathcal{A}$ — may be discrete or continuous (e.g., turn angle, speed)
- Policy: $\pi_\theta : \mathcal{S} \to \mathcal{A}$ — we want to learn the policy parameters θ
- Optimal action: $a^* \in \mathcal{A}$ — provided by expert demonstrator
- Optimal policy: $\pi^* : \mathcal{S} \to \mathcal{A}$ — provided by expert demonstrator
- State dynamics: $P(s_{i+1}|s_i, a_i)$ — simulator, typically not known to policy
  Often deterministic: $s_{i+1} = T(s_i, a_i)$ — deterministic mapping
- Rollout: Given s_0, sequentially execute $a_i = \pi_\theta(s_i)$ & sample $s_{i+1} \sim P(s_{i+1}|s_i, a_i)$
  yields trajectory $\tau = (s_0, a_0, s_1, a_1, \dots)$
- Loss function: $\mathcal{L}(a^*, a)$ — loss of action a given optimal action a^*

Imitation Learning

Components:

- State: $s \in \mathcal{S}$ — may be partially observed (e.g., game screen)
- Action: $a \in \mathcal{A}$ — may be discrete or continuous (e.g., turn angle, speed)
- Policy: $\pi_\theta : \mathcal{S} \rightarrow \mathcal{A}$ — we want to learn the policy parameters θ
- Optimal action: $a^* \in \mathcal{A}$ — provided by expert demonstrator
- Optimal policy: $\pi^* : \mathcal{S} \rightarrow \mathcal{A}$ — provided by expert demonstrator
- State dynamics: $P(s_{i+1}|s_i, a_i)$ — simulator, typically not known to policy

  Often deterministic: $s_{i+1} = T(s_i, a_i)$ — deterministic mapping
- Rollout: Given s_0, sequentially execute $a_i = \pi_\theta(s_i)$ & sample $s_{i+1} \sim P(s_{i+1}|s_i, a_i)$

  yields trajectory $\tau = (s_0, a_0, s_1, a_1, \dots)$
- Loss function: $\mathcal{L}(a^*, a)$ — loss of action a given optimal action a^*

1.1
Network Design

a) Load demonstrations

```python
def load_demonstrations(data_folder):
    """
    1.1 a)
    Given the folder containing the expert demonstrations, the data gets loaded and
    stored it in two lists: observations and actions.
                    N = number of (observation, action) - pairs
    data_folder:    python string, the path to the folder containing the
                    observation_%05d.npy and action_%05d.npy files
    return:
    observations:   python list of N numpy.ndarrays of size (96, 96, 3)
    actions:        python list of N numpy.ndarrays of size 3
    """
    pass
```

b) Understand training

```
def train(data_folder, trained_network_file):
    """
    Function for training the network.
    """
    infer_action = ClassificationNetwork()
    optimizer = torch.optim.Adam(infer_action.parameters(), lr=1e-2)
    loss_function = nn.CrossEntropyLoss()

    observations, actions = load_demonstrations(data_folder)
    observations = [torch.Tensor(observation) for observation in observations]
    actions = [torch.Tensor(action) for action in actions]

    batches = [batch for batch in zip(observations,
                                      infer_action.actions_to_classes(actions))]

    # setting device on GPU if available, else CPU
    device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')

    nr_epochs = 50
    batch_size = 64
    start_time = time.time()

    for epoch in range(nr_epochs):
        random.shuffle(batches)
```

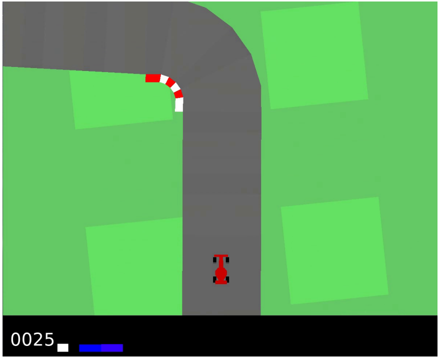
c) Classification Network

- expert demonstrations

  action = [steer, gas, brake]

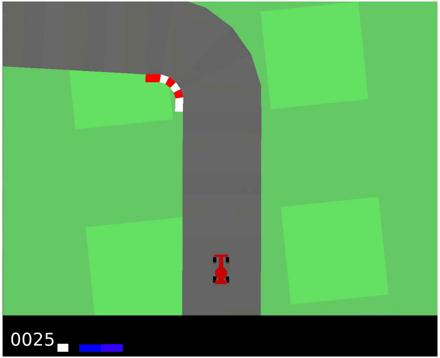
  e.g. `[1., 0., 0.8]`
- define **action-classes**
  - {steer_left}
  - {}
  - {steer_right, brake}
  - {gas}

$\Rightarrow$ map `[1., 0., 0.8]` $\rightarrow$ `?`



c) Classification Network

- `actions_to_classes`
  expert demonstrations $\rightarrow$ action-classes
- `scores_to_action`
  score predicted by the network $\rightarrow$ action
- `CrossEntropyLoss`
  loss function: gt vs. prediction



d) Implement network

```python
class ClassificationNetwork(torch.nn.Module):
    def __init__(self):
        """
        1.1 d)
        Implementation of the network layers. The image size of the input
        observations is 96x96 pixels.
        """
        super().__init__()

        # setting device on GPU if available, else CPU
        device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')

    def forward(self, observation):
        """
        1.1 e)
        The forward pass of the network. Returns the prediction for the given
        input observation.
        observation:   torch.Tensor of size (batch_size, 96, 96, 3)
        return         torch.Tensor of size (batch_size, C)
        """
        pass
```

- 2 to 3 convolution layers + ReLU
- 2 to 3 fully connected layers + ReLU

e) Forward pass, train and test

- color channels or gray-scale
- `python main.py --train`
- `python main.py --test`
- hyper-parameter tuning

e) Forward pass, train and test

```
python main.py --test
```

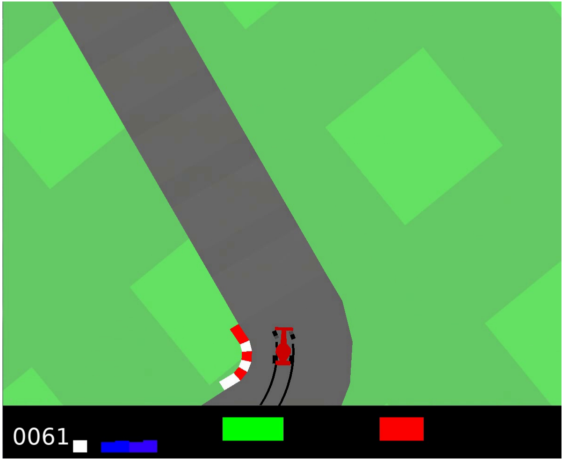
```python
def evaluate(trained_network_file):
    """
    """
    infer_action = torch.load(trained_network_file)
    infer_action.eval()
    env = gym.make('CarRacing-v0')

    # setting device on GPU if available, else CPU
    device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')

    infer_action = infer_action.to(device)


    for episode in range(5):
        observation = env.reset()

        reward_per_episode = 0
        for t in range(500):
            env.render()
            action_scores = infer_action(torch.Tensor(
                np.ascontiguousarray(observation[None])).to(device))

            steer, gas, brake = infer_action.scores_to_action(action_scores)
            observation, reward, done, info = env.step([steer, gas, brake])
            reward_per_episode += reward

        print('episode %d \t reward %f' % (episode, reward_per_episode))
```

f) Record own demonstrations

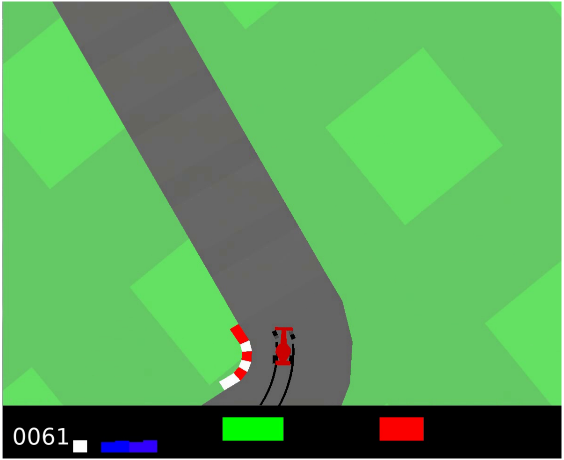
```
def record_demonstrations(demonstrations_folder):
    """
    Function to record own demonstrations by driving the car in the gym car-racing
    environment.
    demonstrations_folder:  python string, the path to where the recorded demonstrations
                            are to be saved

    The controls are:
    arrow keys:             control the car; steer left, steer right, gas, brake
    ESC:                    quit and close
    SPACE:                  restart on a new track
    TAB:                    save the current run
    """
    env = gym.make('CarRacing-v0').env
    status = ControlStatus()
```

f) Record own demonstrations

```
python main.py --teach
```



```
python main.py --test
```



1.2
Network Improvements

Competition

Competition

```
python main.py --score
```

```python
def calculate_score_for_leaderboard(trained_network_file):
    """
    Evaluate the performance of the network. This is the function to be used for
    the final ranking on the course-wide leader-board, only with a different set
    of seeds. Better not change it.
    """
    infer_action = torch.load(trained_network_file)
    infer_action.eval()
    env = gym.make('CarRacing-v0')

    # setting device on GPU if available, else CPU
    device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')

    seeds = [22597174, 68545857, 75568192, 91140053, 86018367,
             49636746, 66759182, 91294619, 84274995, 31531469]

    total_reward = 0

    for episode in range(10):
        env.seed(seeds[episode])
        observation = env.reset()
```

Questions?