2.2C

March 28, 2021

[6]:
```
# install required system dependencies
!apt-get install -y xvfb x11-utils
!apt-get install x11-utils > /dev/null 2>&1
!pip install PyOpenGL==3.1.* \
            PyOpenGL-accelerate==3.1.* \
            gym[box2d]==0.17.*
!pip install pyglet
!pip install ffmpeg
! pip install pyvirtualdisplay
!pip install Image
!pip install gym-maze-trustycoder83
```

```
Reading package lists… Done
Building dependency tree
Reading state information… Done
x11-utils is already the newest version (7.7+3build1).
xvfb is already the newest version (2:1.19.6-1ubuntu4.8).
0 upgraded, 0 newly installed, 0 to remove and 30 not upgraded.
Requirement already satisfied: PyOpenGL==3.1.* in /usr/local/lib/python3.7/dist-
packages (3.1.5)
Requirement already satisfied: PyOpenGL-accelerate==3.1.* in
/usr/local/lib/python3.7/dist-packages (3.1.5)
Requirement already satisfied: gym[box2d]==0.17.* in
/usr/local/lib/python3.7/dist-packages (0.17.2)
Requirement already satisfied: pyglet<=1.5.0,>=1.4.0 in
/usr/local/lib/python3.7/dist-packages (from gym[box2d]==0.17.*) (1.5.0)
Requirement already satisfied: numpy>=1.10.4 in /usr/local/lib/python3.7/dist-
packages (from gym[box2d]==0.17.*) (1.18.5)
Requirement already satisfied: cloudpickle<1.4.0,>=1.2.0 in
/usr/local/lib/python3.7/dist-packages (from gym[box2d]==0.17.*) (1.3.0)
Requirement already satisfied: scipy in /usr/local/lib/python3.7/dist-packages
(from gym[box2d]==0.17.*) (1.4.1)
Requirement already satisfied: box2d-py~=2.3.5; extra == "box2d" in
/usr/local/lib/python3.7/dist-packages (from gym[box2d]==0.17.*) (2.3.8)
Requirement already satisfied: future in /usr/local/lib/python3.7/dist-packages
(from pyglet<=1.5.0,>=1.4.0->gym[box2d]==0.17.*) (0.16.0)
Requirement already satisfied: pyglet in /usr/local/lib/python3.7/dist-packages
(1.5.0)
```

```
Requirement already satisfied: future in /usr/local/lib/python3.7/dist-packages
(from pyglet) (0.16.0)
Requirement already satisfied: ffmpeg in /usr/local/lib/python3.7/dist-packages
(1.4)
Requirement already satisfied: pyvirtualdisplay in
/usr/local/lib/python3.7/dist-packages (2.1)
Requirement already satisfied: EasyProcess in /usr/local/lib/python3.7/dist-
packages (from pyvirtualdisplay) (0.3)
Requirement already satisfied: Image in /usr/local/lib/python3.7/dist-packages
(1.5.33)
Requirement already satisfied: django in /usr/local/lib/python3.7/dist-packages
(from Image) (3.1.7)
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages
(from Image) (1.15.0)
Requirement already satisfied: pillow in /usr/local/lib/python3.7/dist-packages
(from Image) (7.0.0)
Requirement already satisfied: sqlparse>=0.2.2 in /usr/local/lib/python3.7/dist-
packages (from django->Image) (0.4.1)
Requirement already satisfied: asgiref<4,>=3.2.10 in
/usr/local/lib/python3.7/dist-packages (from django->Image) (3.3.1)
Requirement already satisfied: pytz in /usr/local/lib/python3.7/dist-packages
(from django->Image) (2018.9)
Requirement already satisfied: gym-maze-trustycoder83 in
/usr/local/lib/python3.7/dist-packages (0.0.4)
Requirement already satisfied: pygame==1.9.6 in /usr/local/lib/python3.7/dist-
packages (from gym-maze-trustycoder83) (1.9.6)
Requirement already satisfied: numpy==1.18.5 in /usr/local/lib/python3.7/dist-
packages (from gym-maze-trustycoder83) (1.18.5)
Requirement already satisfied: gym==0.17.2 in /usr/local/lib/python3.7/dist-
packages (from gym-maze-trustycoder83) (0.17.2)
Requirement already satisfied: cloudpickle<1.4.0,>=1.2.0 in
/usr/local/lib/python3.7/dist-packages (from gym==0.17.2->gym-maze-
trustycoder83) (1.3.0)
Requirement already satisfied: pyglet<=1.5.0,>=1.4.0 in
/usr/local/lib/python3.7/dist-packages (from gym==0.17.2->gym-maze-
trustycoder83) (1.5.0)
Requirement already satisfied: scipy in /usr/local/lib/python3.7/dist-packages
(from gym==0.17.2->gym-maze-trustycoder83) (1.4.1)
Requirement already satisfied: future in /usr/local/lib/python3.7/dist-packages
(from pyglet<=1.5.0,>=1.4.0->gym==0.17.2->gym-maze-trustycoder83) (0.16.0)
```

We now proceed to initialise the monitor wrapper for Gym so we can visualise the maze and the agent on a video

[12]:
```
import sys
# import pygame
import numpy as np
# import math
```

```
# import base64
# import io
# import IPython
import gym
import gym_maze

# from gym.wrappers import Monitor
# from IPython import display
from pyvirtualdisplay import Display
from gym.wrappers.monitoring import video_recorder

d = Display()
d.start()

# Recording filename
video_name = "./vid/Practical_2.mp4"

# Setup the environment for the maze
env = gym.make("maze-sample-10x10-v0")
```

We now proceed to perform a simple Q-tracking algorithm. The method here is quite sub-optimal since it employs a random choice to pick between exploration and exploitation. It does illustrate, however, how Q-value tracking can be done using the maze.

[13]:
```
def get_Q_table_ind (size, x, y):
  x_max, y_max = size
  return y_max * x + y

def policy (Q_table, current_state, epsilon):
  current_state_in_Q = get_Q_table_ind(SIZE, current_state[0], current_state[1])
  if np.random.uniform(0,1) < epsilon:
      return env.action_space.sample()
  else:
      return int(np.argmax(Q_table[current_state_in_Q]))

def Q_update (Q_table, current_state, next_state, action, reward,␣
 ↪learning_rate):
  current_state_in_Q = get_Q_table_ind(SIZE, current_state[0], current_state[1])
  next_state_in_Q = get_Q_table_ind(SIZE, next_state[0], next_state[1])
  Q_table[current_state_in_Q, action] = (1-learning_rate)␣
 ↪*Q_table[current_state_in_Q, action] +learning_rate*(reward +␣
 ↪max(Q_table[next_state_in_Q,:]))

def Q_update_new (Q_table, current_state, next_state, action, reward, alpha,␣
 ↪gamma):
  current_state_in_Q = get_Q_table_ind(SIZE, current_state[0], current_state[1])
  next_state_in_Q = get_Q_table_ind(SIZE, next_state[0], next_state[1])
```

```
  Q_table[current_state_in_Q, action] = Q_table[current_state_in_Q, action] +␣
↪alpha * (reward - gamma * Q_table[current_state_in_Q, action])
```

[9]:
```
def run(env, epsilon, gamma):
  n_actions = env.action_space.n
  current_state = env.reset()
  Q_table = np.zeros((XSIZE * YSIZE, n_actions))
  N_count = np.zeros(XSIZE * YSIZE)
  rewards = []

  for e in range(50):
      total_reward = 0
      # We are not done yet
      done = False
      for i in range(MAX_ITERATION):
          env.unwrapped.render()
          vid.capture_frame()

          action = policy(Q_table, current_state, epsilon)
          next_state, reward, done, _ = env.step(action)

          # Q_update(Q_table, current_state, next_state, action, reward,␣
↪learning_rate)
          current_state_in_Q = get_Q_table_ind(SIZE, current_state[0],␣
↪current_state[1])
          N_count[current_state_in_Q] += 1
          Q_update_new (Q_table, current_state, next_state, action, reward, 1.0␣
↪/N_count[current_state_in_Q], gamma)
          total_reward = total_reward + reward

          # If the episode is finished, we leave the for loop
          if done:
              break
          current_state = next_state

      #Show reward
      print("Total episode reward:", total_reward)
      rewards.append(total_reward)
      current_state = env.reset()
  return rewards
```

[14]:
```
MAX_ITERATION = 500
XSIZE = 10
YSIZE = 10
SIZE = [XSIZE, YSIZE]

# Setup the video
```

```
vid = None
vid = video_recorder.VideoRecorder(env,video_name)

epsilon = 0
gamma = 0.5

rewards_0_05 = run(env, epsilon, gamma)

print("Video successfuly saved.")
vid.close()
vid.enabled = False
```

```
Total episode reward: -0.5000000000000003
Total episode reward: 0.7119999999999997
Total episode reward: 0.7389999999999999
Total episode reward: 0.6969999999999998
Total episode reward: 0.6589999999999998
Total episode reward: 0.6479999999999997
Total episode reward: 0.6449999999999998
Total episode reward: 0.6159999999999997
Total episode reward: 0.5359999999999996
Total episode reward: -0.5000000000000003
Total episode reward: 0.5529999999999997
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.6889999999999998
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.6679999999999997
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.7519999999999998
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
```

```
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.6519999999999997
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.5889999999999997
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Video successfuly saved.
```

[17]:
```
epsilon = 0.1
gamma = 0.5
rewards_01_05 = run(env, epsilon, gamma)
```

```
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.6209999999999998
Total episode reward: -0.5000000000000003
Total episode reward: 0.7849999999999998
Total episode reward: -0.5000000000000003
Total episode reward: 0.6819999999999997
Total episode reward: 0.6109999999999998
Total episode reward: 0.5559999999999996
Total episode reward: 0.5799999999999996
Total episode reward: 0.5699999999999996
Total episode reward: 0.7079999999999997
Total episode reward: -0.5000000000000003
Total episode reward: 0.5859999999999996
Total episode reward: 0.5319999999999996
Total episode reward: 0.5739999999999996
Total episode reward: -0.5000000000000003
Total episode reward: 0.5579999999999996
Total episode reward: -0.5000000000000003
Total episode reward: 0.7009999999999998
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.6119999999999997
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.6659999999999997
```

```
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.7529999999999998
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.6459999999999997
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.6259999999999997
Total episode reward: -0.5000000000000003
```

[ ]:
```
epsilon = 0.5
gamma = 0.5
rewards_05_05 = run(env, epsilon, gamma)
```

```
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.5889999999999997
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
```

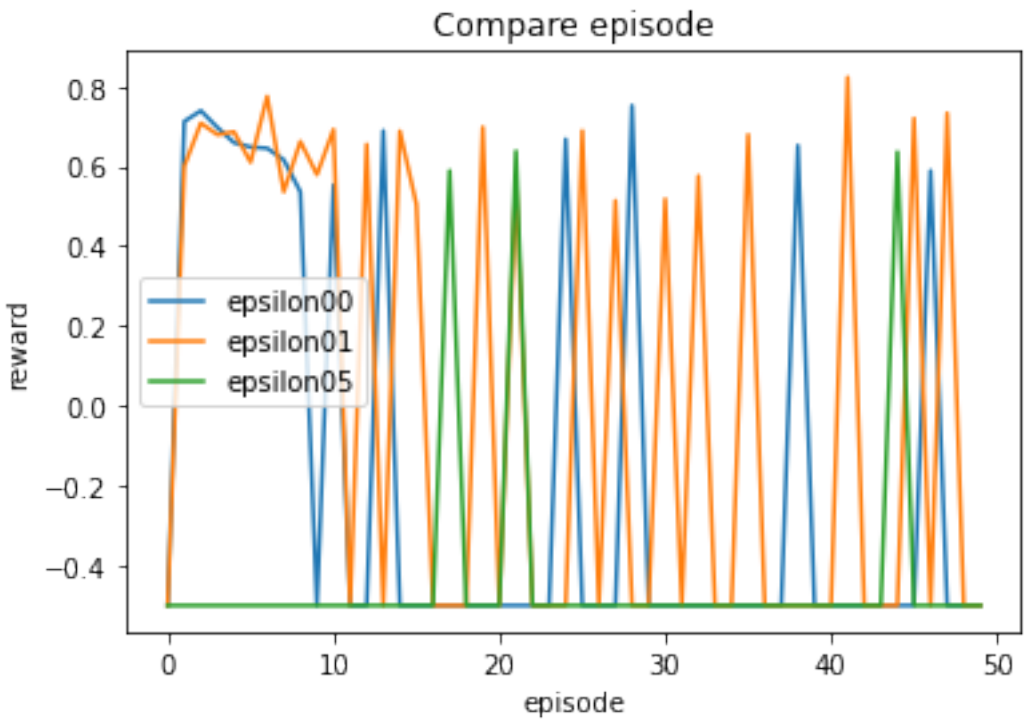
```
Total episode reward: -0.5000000000000003
Total episode reward: 0.6379999999999997
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.6359999999999997
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
```

[ ]:
```python
import matplotlib.pyplot as plt
plt.plot(rewards_0_05, label = "epsilon00")
plt.plot(rewards_01_05, label = "epsilon01")
plt.plot(rewards_05_05, label = "epsilon05")
plt.xlabel('episode')
plt.ylabel('reward')

plt.title('Compare episode')

plt.legend()

plt.show()
```



[15]:
```
epsilon = 0.1
gamma = 0.0
rewards_01_00 = run(env, epsilon, gamma)
```

```
Total episode reward: 0.6839999999999997
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.7199999999999998
Total episode reward: 0.5279999999999996
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.7669999999999998
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.6479999999999997
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
```

```
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.7639999999999998
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.5179999999999996
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.7459999999999998
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.7729999999999998
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
```

[16]:
```
epsilon = 0.1
gamma = 1.0
rewards_01_10 = run(env, epsilon, gamma)
```

```
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.7999999999999998
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.5549999999999997
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
```

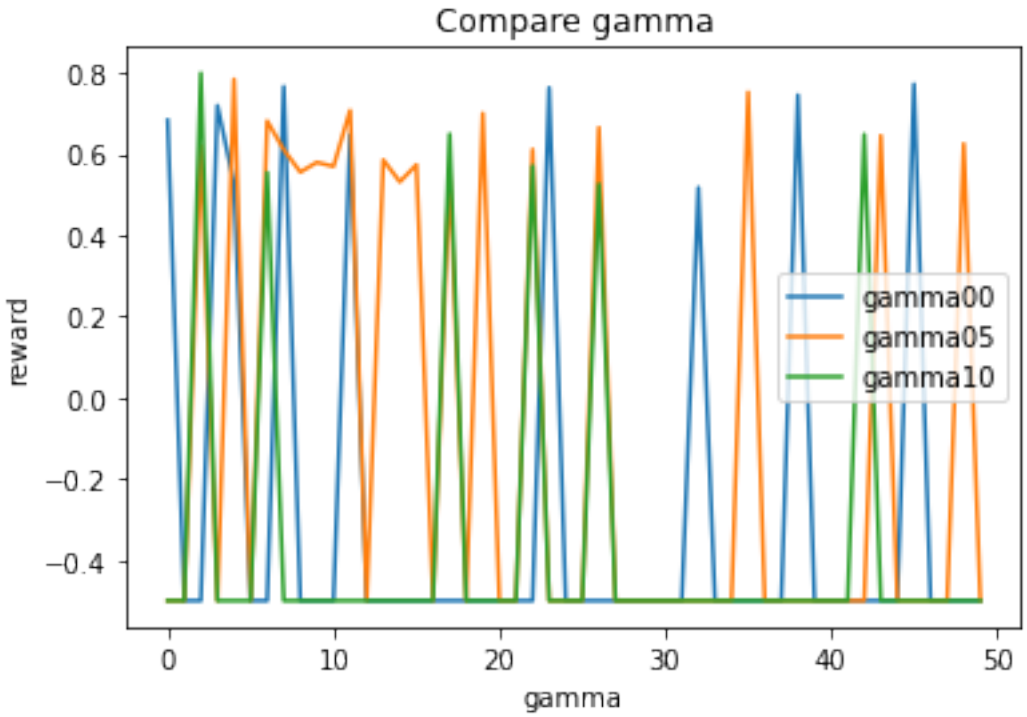
```
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.6499999999999997
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.5709999999999997
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.5259999999999996
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: 0.6489999999999998
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
Total episode reward: -0.5000000000000003
```

[18]:
```
import matplotlib.pyplot as plt
plt.plot(rewards_01_00, label = "gamma00")
plt.plot(rewards_01_05, label = "gamma05")
plt.plot(rewards_01_10, label = "gamma10")
plt.xlabel('gamma')
plt.ylabel('reward')

plt.title('Compare gamma')
```

```
plt.legend()

plt.show()
```



[ ]:
```
import base64
import io
from IPython import display

video_name = "./vid/Practical_2.mp4"

video = io.open(video_name, 'r+b').read()
encoded = base64.b64encode(video)

display.display(display.HTML(data="""
  <video alt="test" controls>
  <source src="data:video/mp4;base64,{0}" type="video/mp4" />
  </video>
  """.format(encoded.decode('ascii'))))
```

```
<IPython.core.display.HTML object>
```