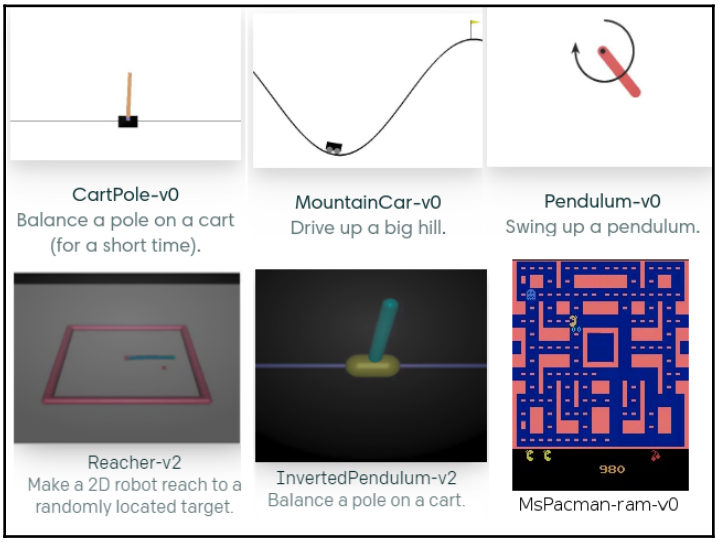
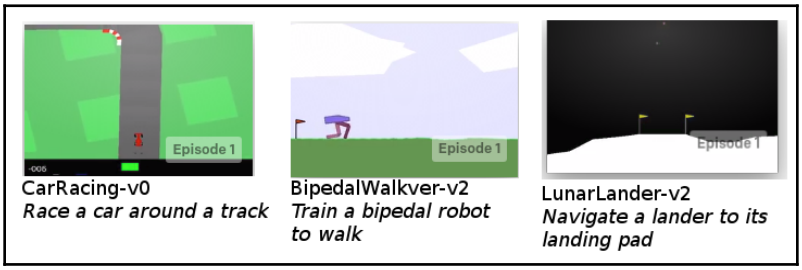
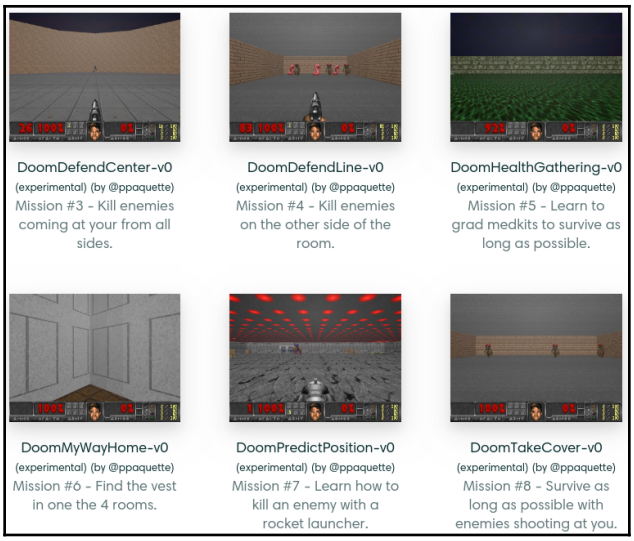
Chapter 1: Introduction to Intelligent Agents and Learning Environments







DoomDefendCenter-v0
(experimental) (by @ppaquette)
Mission #3 - Kill enemies coming at your from all sides.

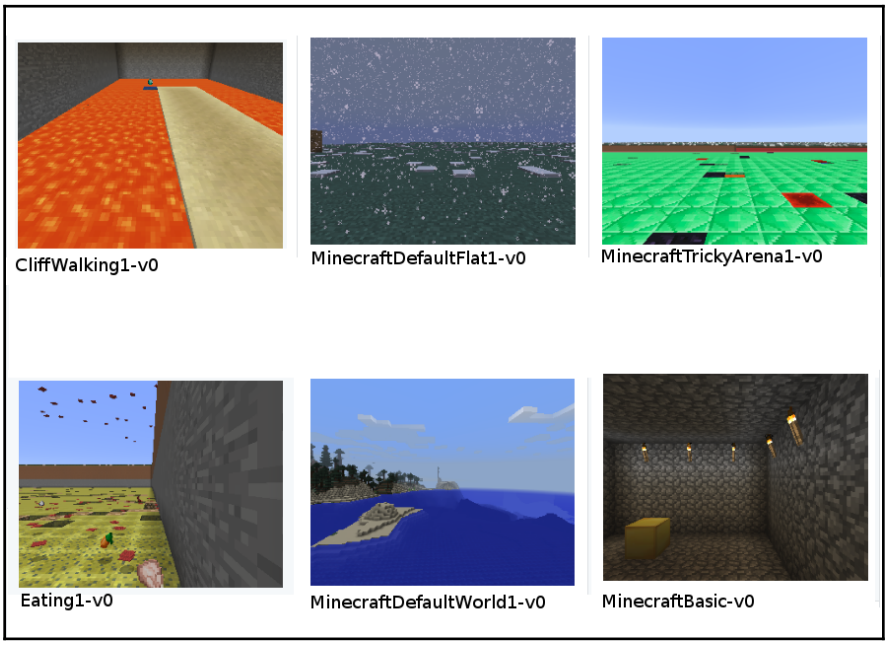
DoomDefendLine-v0
(experimental) (by @ppaquette)
Mission #4 - Kill enemies on the other side of the room.

DoomHealthGathering-v0
(experimental) (by @ppaquette)
Mission #5 - Learn to grad medkits to survive as long as possible.

DoomMyWayHome-v0
(experimental) (by @ppaquette)
Mission #6 - Find the vest in one the 4 rooms.

DoomPredictPosition-v0
(experimental) (by @ppaquette)
Mission #7 - Learn how to kill an enemy with a rocket launcher.

DoomTakeCover-v0
(experimental) (by @ppaquette)
Mission #8 - Survive as long as possible with enemies shooting at you.



CliffWalking1-v0

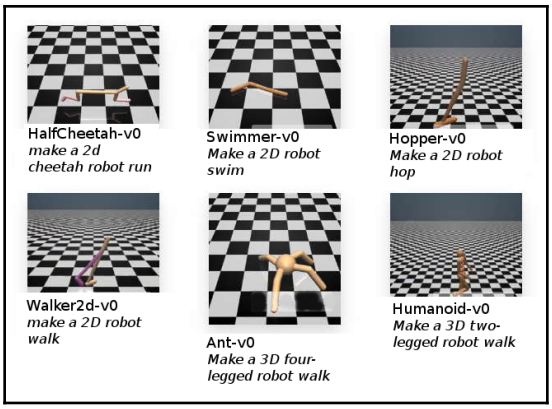
MinecraftDefaultFlat1-v0

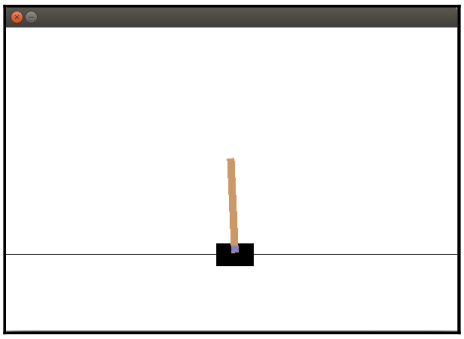
MinecraftTrickyArena1-v0

Eating1-v0

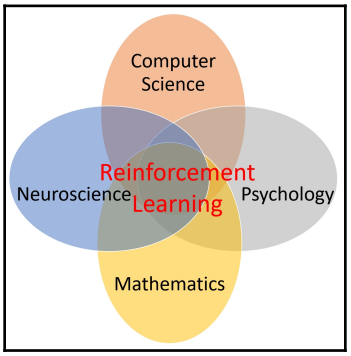
MinecraftDefaultWorld1-v0

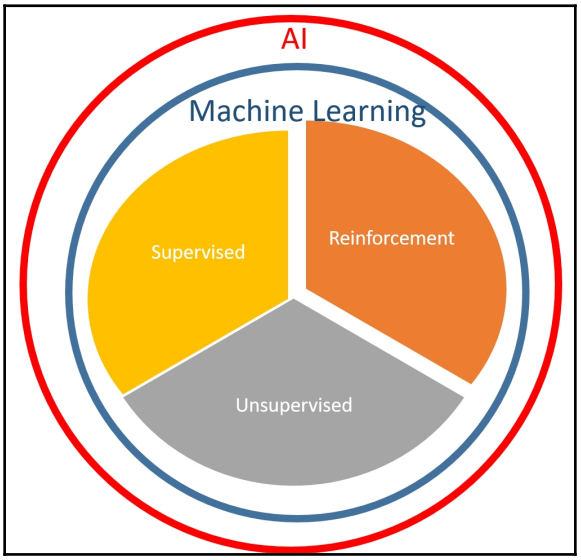
MinecraftBasic-v0

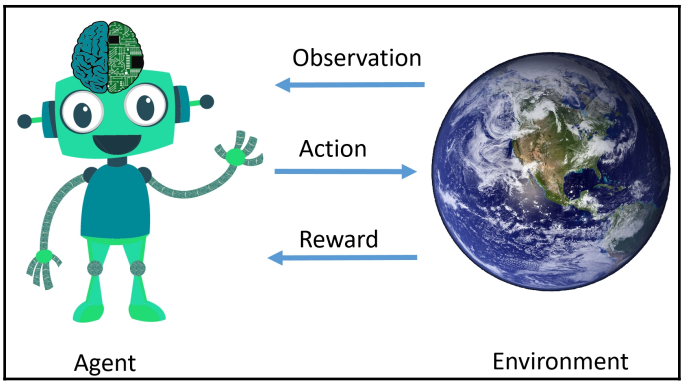

HalfCheetah-v0
make a 2d cheetah robot run
Swimmer-v0
Make a 2D robot swim
Hopper-v0
Make a 2D robot hop
Walker2d-v0
make a 2D robot walk
Ant-v0
Make a 3D four-legged robot walk
Humanoid-v0
Make a 3D two-legged robot walk

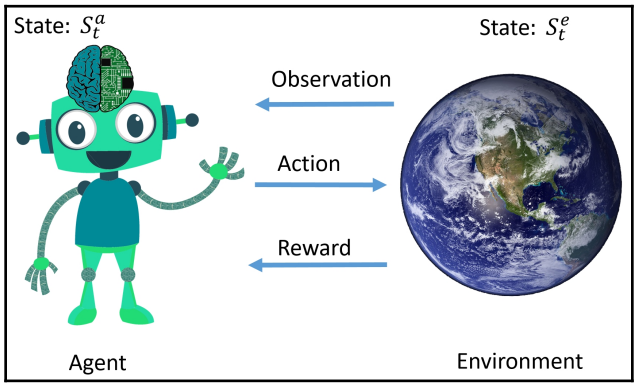


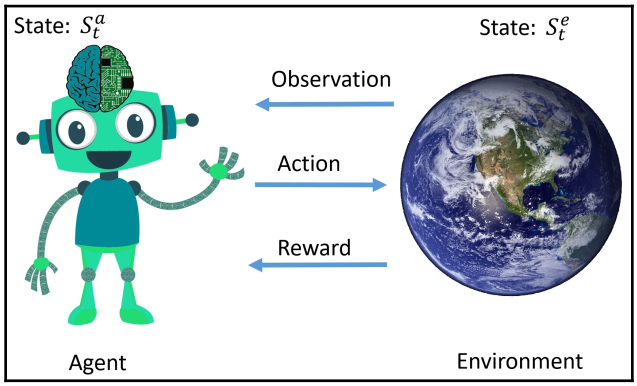
Chapter 2: Reinforcement Learning and Deep Reinforcement Learning





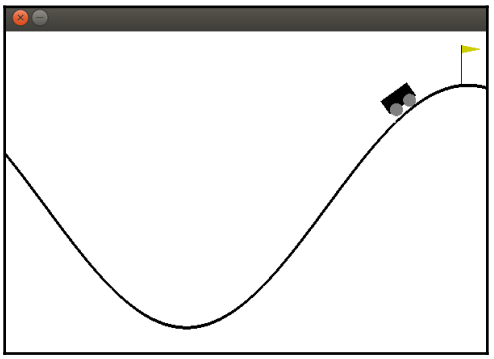

Observation
Action
Reward
Agent
Environment


State: S_t^a
State: S_t^e
Observation
Action
Reward
Agent
Environment

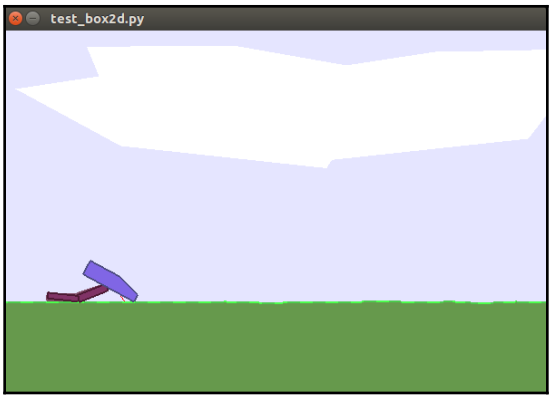

State: S_t^a
State: S_t^e
Observation
Action
Reward
Agent
Environment


Markov Process
Markov Reward Process
Markov Decision Process

Chapter 3: Getting Started with OpenAI Gym and Deep Reinforcement Learning



```
(array([-0.53439364,  0.05070148]), -1.0, True, {})
(array([-0.48461121,  0.04978243]), -1.0, True, {})
(array([-0.43512052,  0.04949069]), -1.0, True, {})
(array([-0.38728565,  0.04783487]), -1.0, True, {})
(array([-0.33944487,  0.04784078]), -1.0, True, {})
(array([-0.29291606,  0.04652882]), -1.0, True, {})
(array([-0.24898253,  0.04393353]), -1.0, True, {})
(array([-0.20688342,  0.04209911]), -1.0, True, {})
(array([-0.16781806,  0.03906536]), -1.0, True, {})
(array([-0.1299425 ,  0.03787556]), -1.0, True, {})
(array([-0.09337938,  0.03656312]), -1.0, True, {})
(array([-0.05921881,  0.03416058]), -1.0, True, {})
(array([-0.02751888,  0.03169993]), -1.0, True, {})
(array([0.00068956, 0.02820844]), -1.0, True, {})
(array([0.02639801, 0.02570845]), -1.0, True, {})
(array([0.05061429, 0.02421628]), -1.0, True, {})
(array([0.07335933, 0.02274505]), -1.0, True, {})
(array([0.09466468, 0.02130535]), -1.0, True, {})
(array([0.11357016, 0.01890549]), -1.0, True, {})
(array([0.12911936, 0.01554919]), -1.0, True, {})
(array([0.14335377, 0.01423442]), -1.0, True, {})
(array([0.15631584, 0.01296207]), -1.0, True, {})
>>>
```

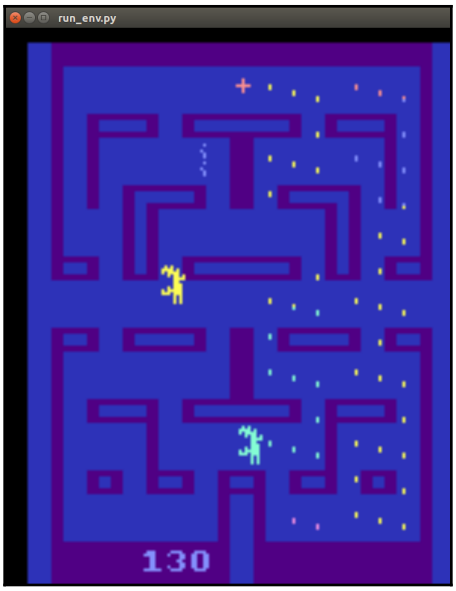


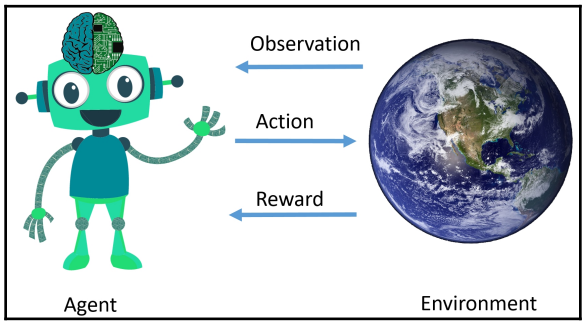

```
praveen@ubuntu: ~
praveen@ubuntu:~$ nvcc -V
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2016 NVIDIA Corporation
Built on Sun_Sep__4_22:14:01_CDT_2016
Cuda compilation tools, release 8.0, V8.0.44
praveen@ubuntu:~$
```

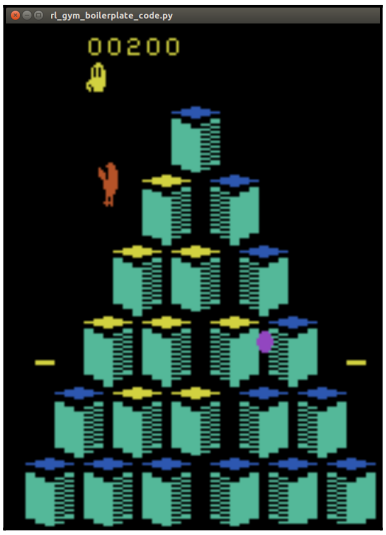
```
praveen@ubuntu: ~/rl_gym_book
(rl_gym_book) praveen@ubuntu:~/rl_gym_book$ python pytorch_test.py

1.00000e-36 *
  2.9928  0.0000  0.0000
  0.0000  0.1263  0.0000
  0.1326  0.0000  0.0000
[torch.FloatTensor of size 3x3]

(rl_gym_book) praveen@ubuntu:~/rl_gym_book$
```

Chapter 4: Exploring the Gym and its Features

```
(rl_gym_book) praveen@ubuntu:~/rl_gym_book/ch4$ python list_gym_envs.py
Acrobot-v1
AirRaid-ram-v0
AirRaid-ram-v4
AirRaid-ramDeterministic-v0
AirRaid-ramDeterministic-v4
AirRaid-ramNoFrameskip-v0
AirRaid-ramNoFrameskip-v4
AirRaid-v0
AirRaid-v4
AirRaidDeterministic-v0
AirRaidDeterministic-v4
AirRaidNoFrameskip-v0
AirRaidNoFrameskip-v4
Alien-ram-v0
Alien-ram-v4
Alien-ramDeterministic-v0
Alien-ramDeterministic-v4
Alien-ramNoFrameskip-v0
Alien-ramNoFrameskip-v4
Alien-v0
Alien-v4
AlienDeterministic-v0
AlienDeterministic-v4
AlienNoFrameskip-v0
AlienNoFrameskip-v4
Amidar-ram-v0
Amidar-ram-v4
Amidar-ramDeterministic-v0
Amidar-ramDeterministic-v4
Amidar-ramNoFrameskip-v0
Amidar-ramNoFrameskip-v4
Amidar-v0
Amidar-v4
AmidarDeterministic-v0
```


run_env.py
130


Observation
Action
Reward
Agent
Environment

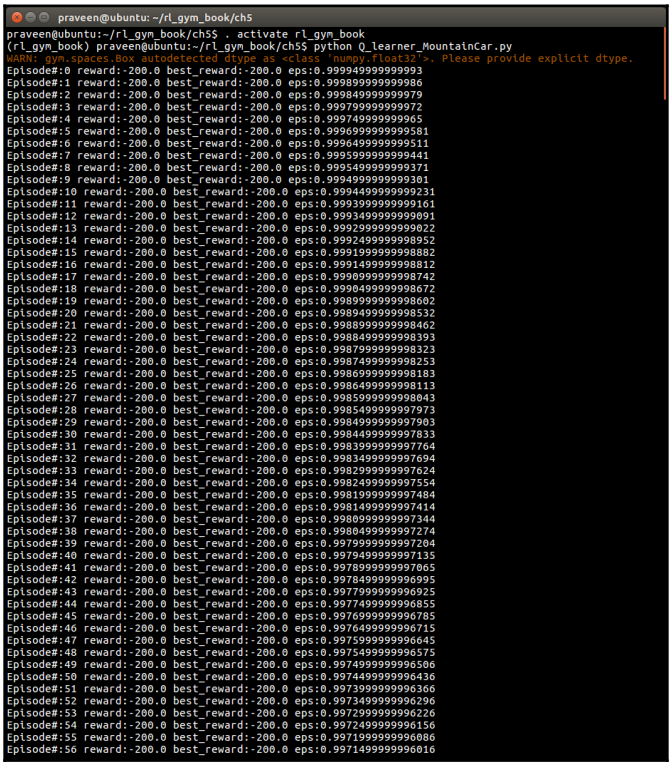


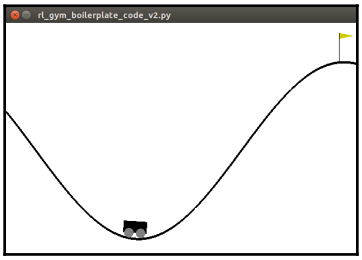

```
(rl_gym_book) praveen@ubuntu:~/rl_gym_book/ch4$ python rl_gym_boilerplate_code.py

 Episode #0 ended in 375 steps.

 Episode #1 ended in 363 steps.

 Episode #3 ended in 495 steps.

 Episode #4 ended in 437 steps.

 Episode #5 ended in 355 steps.

 Episode #6 ended in 443 steps.

 Episode #7 ended in 407 steps.

 Episode #8 ended in 400 steps.

 Episode #9 ended in 376 steps.
```

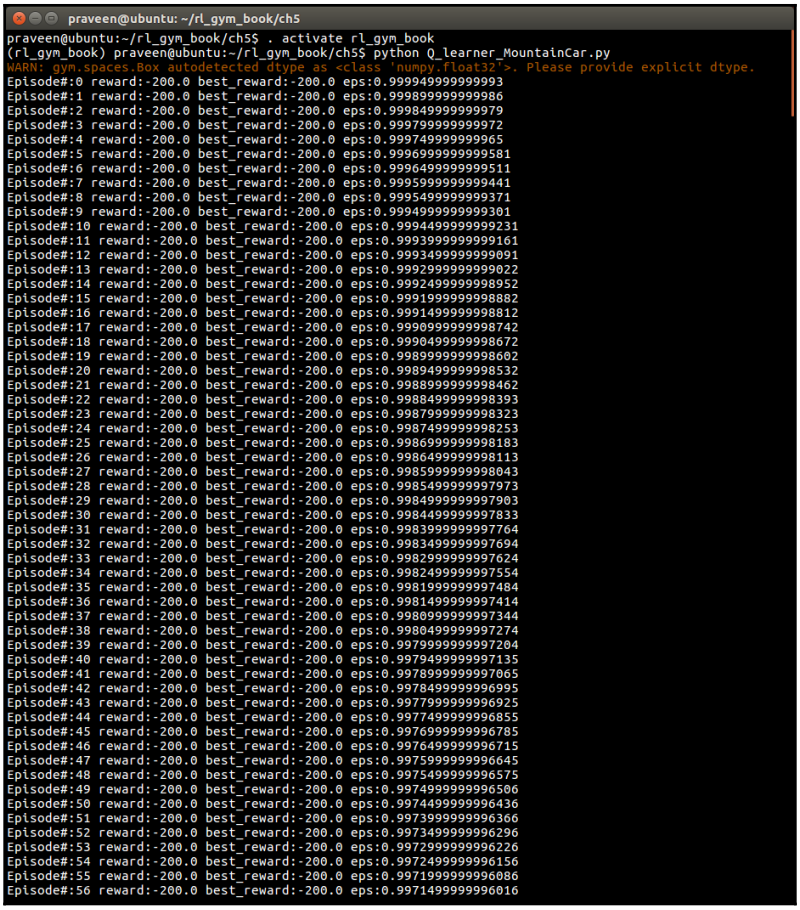
```
(rl_gym_book) praveen@ubuntu:~/rl_gym_book/ch4$ python get_observation_action_space.py CartPole-v0
WARN: gym.spaces.Box autodetected dtype as <class 'numpy.float32'>. Please provide explicit dtype.
Observation Space:
Box(4,)

 space.low:  [ -4.80000019e+00  -3.40282347e+38  -4.18879032e-01  -3.40282347e+38]

 space.high:  [  4.80000019e+00   3.40282347e+38   4.18879032e-01   3.40282347e+38]
Action Space:
Discrete(2)
```

```
(rl_gym_book) praveen@ubuntu:~/rl_gym_book/ch4$ python get_observation_action_space.py BipedalWalker-v2
WARN: gym.spaces.Box autodetected dtype as <class 'numpy.float32'>. Please provide explicit dtype.
WARN: gym.spaces.Box autodetected dtype as <class 'numpy.float32'>. Please provide explicit dtype.
Observation Space:
Box(24,)

 space.low:  [-inf -inf -inf -inf -inf -inf -inf -inf -inf -inf -inf -inf -inf -inf
 -inf -inf -inf -inf -inf -inf -inf -inf -inf]

 space.high:  [ inf  inf  inf  inf  inf  inf  inf  inf  inf  inf  inf  inf  inf  inf
  inf  inf  inf  inf  inf  inf  inf  inf  inf]
Action Space:
Box(4,)

 space.low:  [-1. -1. -1. -1.]

 space.high:  [ 1.  1.  1.  1.]
```

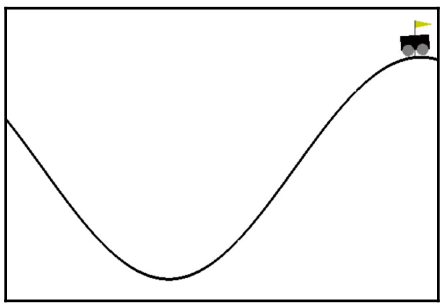
Chapter 5: Implementing your First Learning Agent - Solving the Mountain Car problem



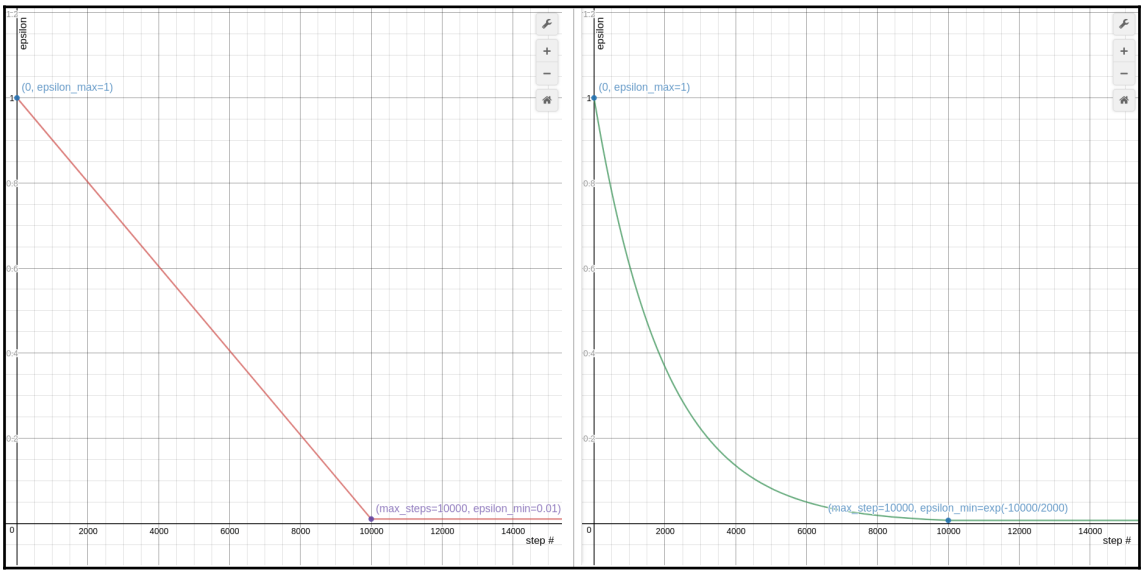



praveen@ubuntu: ~/rl_gym_book/ch5
praveen@ubuntu:~/rl_gym_book/ch5$. activate rl_gym_book
(rl_gym_book) praveen@ubuntu:~/rl_gym_book/ch5$ python Q_learner_MountainCar.py
WARN: gym.spaces.Box autodetected dtype as <class 'numpy.float32'>. Please provide explicit dtype.
Episode#:0 reward:-200.0 best_reward:-200.0 eps:0.999949999999993
Episode#:1 reward:-200.0 best_reward:-200.0 eps:0.999899999999986
Episode#:2 reward:-200.0 best_reward:-200.0 eps:0.999849999999979
Episode#:3 reward:-200.0 best_reward:-200.0 eps:0.999799999999972
Episode#:4 reward:-200.0 best_reward:-200.0 eps:0.999749999999965
Episode#:5 reward:-200.0 best_reward:-200.0 eps:0.999699999999581
Episode#:6 reward:-200.0 best_reward:-200.0 eps:0.999649999999511
Episode#:7 reward:-200.0 best_reward:-200.0 eps:0.999599999999441
Episode#:8 reward:-200.0 best_reward:-200.0 eps:0.999549999999371
Episode#:9 reward:-200.0 best_reward:-200.0 eps:0.999499999999301
Episode#:10 reward:-200.0 best_reward:-200.0 eps:0.999449999999231
Episode#:11 reward:-200.0 best_reward:-200.0 eps:0.999399999999161
Episode#:12 reward:-200.0 best_reward:-200.0 eps:0.999349999999091
Episode#:13 reward:-200.0 best_reward:-200.0 eps:0.999299999999022
Episode#:14 reward:-200.0 best_reward:-200.0 eps:0.999249999998952
Episode#:15 reward:-200.0 best_reward:-200.0 eps:0.999199999998882
Episode#:16 reward:-200.0 best_reward:-200.0 eps:0.999149999998812
Episode#:17 reward:-200.0 best_reward:-200.0 eps:0.999099999998742
Episode#:18 reward:-200.0 best_reward:-200.0 eps:0.999049999998672
Episode#:19 reward:-200.0 best_reward:-200.0 eps:0.998999999998602
Episode#:20 reward:-200.0 best_reward:-200.0 eps:0.998949999998532
Episode#:21 reward:-200.0 best_reward:-200.0 eps:0.998899999998462
Episode#:22 reward:-200.0 best_reward:-200.0 eps:0.998849999998393
Episode#:23 reward:-200.0 best_reward:-200.0 eps:0.998799999998323
Episode#:24 reward:-200.0 best_reward:-200.0 eps:0.998749999998253
Episode#:25 reward:-200.0 best_reward:-200.0 eps:0.998699999998183
Episode#:26 reward:-200.0 best_reward:-200.0 eps:0.998649999998113
Episode#:27 reward:-200.0 best_reward:-200.0 eps:0.998599999998043
Episode#:28 reward:-200.0 best_reward:-200.0 eps:0.998549999997973
Episode#:29 reward:-200.0 best_reward:-200.0 eps:0.998499999997903
Episode#:30 reward:-200.0 best_reward:-200.0 eps:0.998449999997833
Episode#:31 reward:-200.0 best_reward:-200.0 eps:0.998399999997764
Episode#:32 reward:-200.0 best_reward:-200.0 eps:0.998349999997694
Episode#:33 reward:-200.0 best_reward:-200.0 eps:0.998299999997624
Episode#:34 reward:-200.0 best_reward:-200.0 eps:0.998249999997554
Episode#:35 reward:-200.0 best_reward:-200.0 eps:0.998199999997484
Episode#:36 reward:-200.0 best_reward:-200.0 eps:0.998149999997414
Episode#:37 reward:-200.0 best_reward:-200.0 eps:0.998099999997344
Episode#:38 reward:-200.0 best_reward:-200.0 eps:0.998049999997274
Episode#:39 reward:-200.0 best_reward:-200.0 eps:0.997999999997204
Episode#:40 reward:-200.0 best_reward:-200.0 eps:0.997949999997135
Episode#:41 reward:-200.0 best_reward:-200.0 eps:0.997899999997065
Episode#:42 reward:-200.0 best_reward:-200.0 eps:0.997849999996995
Episode#:43 reward:-200.0 best_reward:-200.0 eps:0.997799999996925
Episode#:44 reward:-200.0 best_reward:-200.0 eps:0.997749999996855
Episode#:45 reward:-200.0 best_reward:-200.0 eps:0.997699999996785
Episode#:46 reward:-200.0 best_reward:-200.0 eps:0.997649999996715
Episode#:47 reward:-200.0 best_reward:-200.0 eps:0.997599999996645
Episode#:48 reward:-200.0 best_reward:-200.0 eps:0.997549999996575
Episode#:49 reward:-200.0 best_reward:-200.0 eps:0.997499999996506
Episode#:50 reward:-200.0 best_reward:-200.0 eps:0.997449999996436
Episode#:51 reward:-200.0 best_reward:-200.0 eps:0.997399999996366
Episode#:52 reward:-200.0 best_reward:-200.0 eps:0.997349999996296
Episode#:53 reward:-200.0 best_reward:-200.0 eps:0.997299999996226
Episode#:54 reward:-200.0 best_reward:-200.0 eps:0.997249999996156
Episode#:55 reward:-200.0 best_reward:-200.0 eps:0.997199999996086
Episode#:56 reward:-200.0 best_reward:-200.0 eps:0.997149999996016

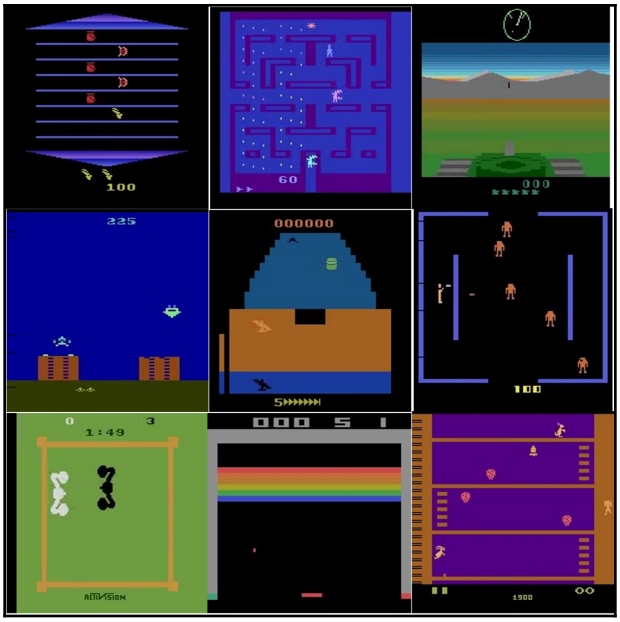
```
praveen@ubuntu: ~/rl_gym_book/ch5
Episode#:34509 reward:-152.0 best_reward:-94.0 eps:0.0049999999486979654
Episode#:34510 reward:-151.0 best_reward:-94.0 eps:0.0049999999486979654
Episode#:34511 reward:-147.0 best_reward:-94.0 eps:0.0049999999486979654
Episode#:34512 reward:-156.0 best_reward:-94.0 eps:0.0049999999486979654
Episode#:34513 reward:-111.0 best_reward:-94.0 eps:0.0049999999486979654
Episode#:34514 reward:-98.0 best_reward:-94.0 eps:0.0049999999486979654
Episode#:34515 reward:-117.0 best_reward:-94.0 eps:0.0049999999486979654
Episode#:34516 reward:-113.0 best_reward:-94.0 eps:0.0049999999486979654
Episode#:34517 reward:-149.0 best_reward:-94.0 eps:0.0049999999486979654
Episode#:34518 reward:-116.0 best_reward:-94.0 eps:0.0049999999486979654
Episode#:34519 reward:-149.0 best_reward:-94.0 eps:0.0049999999486979654
Episode#:34520 reward:-90.0 best_reward:-90.0 eps:0.0049999999486979654
Episode#:34521 reward:-112.0 best_reward:-90.0 eps:0.0049999999486979654
Episode#:34522 reward:-160.0 best_reward:-90.0 eps:0.0049999999486979654
Episode#:34523 reward:-112.0 best_reward:-90.0 eps:0.0049999999486979654
Episode#:34524 reward:-94.0 best_reward:-90.0 eps:0.0049999999486979654
Episode#:34525 reward:-90.0 best_reward:-90.0 eps:0.0049999999486979654
Episode#:34526 reward:-151.0 best_reward:-90.0 eps:0.0049999999486979654
Episode#:34527 reward:-119.0 best_reward:-90.0 eps:0.0049999999486979654
Episode#:34528 reward:-114.0 best_reward:-90.0 eps:0.0049999999486979654
Episode#:34529 reward:-86.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34530 reward:-89.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34531 reward:-117.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34532 reward:-162.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34533 reward:-152.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34534 reward:-108.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34535 reward:-118.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34536 reward:-149.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34537 reward:-114.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34538 reward:-121.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34539 reward:-153.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34540 reward:-89.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34541 reward:-147.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34542 reward:-115.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34543 reward:-150.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34544 reward:-119.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34545 reward:-150.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34546 reward:-113.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34547 reward:-115.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34548 reward:-119.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34549 reward:-158.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34550 reward:-121.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34551 reward:-87.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34552 reward:-115.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34553 reward:-147.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34554 reward:-115.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34555 reward:-120.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34556 reward:-147.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34557 reward:-114.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34558 reward:-165.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34559 reward:-148.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34560 reward:-152.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34561 reward:-162.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34562 reward:-112.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34563 reward:-95.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34564 reward:-149.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34565 reward:-147.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34566 reward:-114.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34567 reward:-92.0 best_reward:-86.0 eps:0.0049999999486979654
Episode#:34568 reward:-93.0 best_reward:-86.0 eps:0.0049999999486979654
```



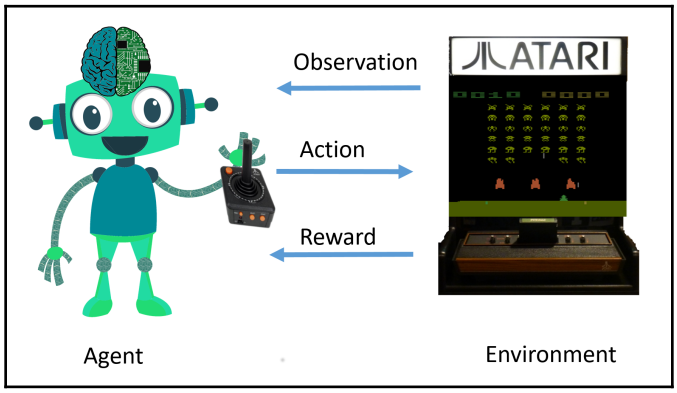
Chapter 6: Implementing an Intelligent Agent for Optimal Control using Deep Q-Learning



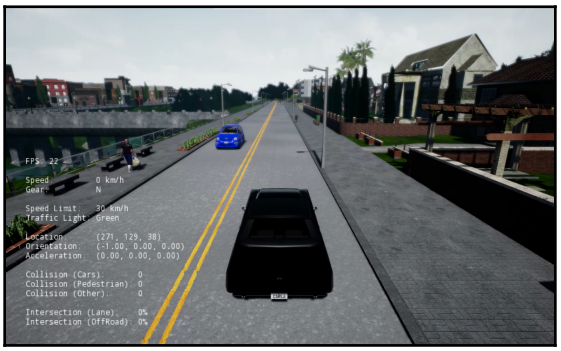

DQL
DQL/epsilon
0.0640
0.0600
0.0560
0.0520
0.0480
0.000 15.00M 30.00M 45.00M
DQL/td_error
0.0220
0.0160
0.0100
4.000e-3
-2.000e-3
5.000M 25.00M 45.00M
main
main/ep_reward
200
160
120
80.0
40.0
0.00
0.000 15.00M 30.00M 45.00M
main/max_ep_rew
500
400
300
200
100
0.00
0.000 15.00M 30.00M 45.00M
main/mean_ep_reward
120
80.0
40.0
0.00
0.000 15.00M 30.00M 45.00M




Atari Gym Observation
Shape: (210 x 160 x 3)
Crop
Shape: ~(160 x 160 x 3)
RGB2GrayScale
Shape: ~(160 x 160)
Down-sample/Resize
Shape: (84 x 84 x 1)
Processed Observation
Shape: (84 x 84)


Observation
Action
Reward
ATARI
Agent
Environment

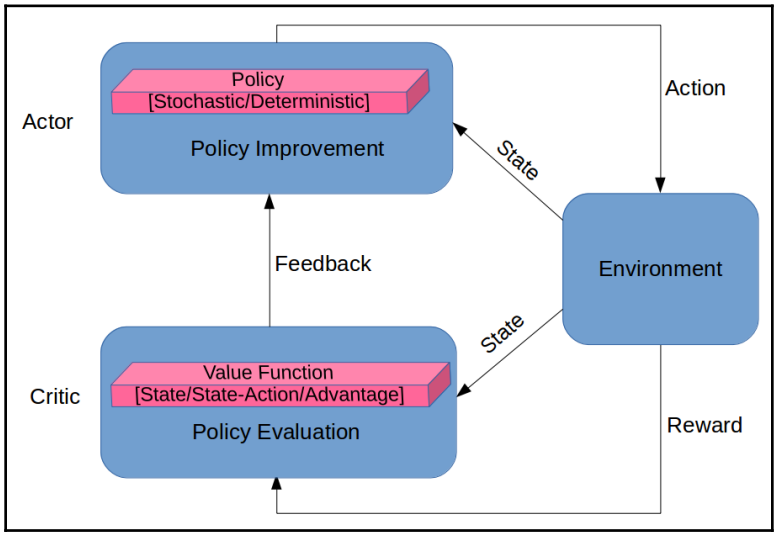
Chapter 7: Creating Custom OpenAI Gym Environments - CARLA Driving Simulator

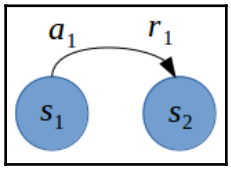


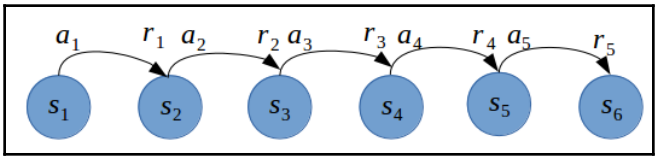

```
praveen@ubuntu: ~/rl_gym_book/ch7
(rl_gym_book) praveen@ubuntu:~/rl_gym_book/ch7$ python carla-gym/carla_gym/envs/
carla_env.py
Initializing new Carla server...
Start pos 36 ([0.0, 3.0]), end 40 ([0.0, 3.0])
Starting new episode...
step#: 1 reward: -0.0 total_reward: -0.0 done: False
step#: 2 reward: -0.0 total_reward: -0.0 done: False
step#: 3 reward: -0.0 total_reward: -0.0 done: False
step#: 4 reward: -0.0 total_reward: -0.0 done: False
step#: 5 reward: -0.0 total_reward: -0.0 done: False
step#: 6 reward: -0.0007 total_reward: -0.0007 done: False
step#: 7 reward: 0.0006 total_reward: -0.0001 done: False
step#: 8 reward: 0.0001 total_reward: 0.0001 done: False
step#: 9 reward: 0.0 total_reward: 0.0001 done: False
step#: 10 reward: -0.0 total_reward: 0.0001 done: False
step#: 11 reward: -0.0 total_reward: 0.0 done: False
step#: 12 reward: -0.0 total_reward: 0.0 done: False
step#: 13 reward: -0.0 total_reward: 0.0 done: False
step#: 14 reward: -0.0 total_reward: 0.0 done: False
step#: 15 reward: -0.0 total_reward: 0.0 done: False
step#: 16 reward: -0.0 total_reward: 0.0 done: False
step#: 17 reward: -0.0 total_reward: 0.0 done: False
step#: 18 reward: -0.0 total_reward: 0.0 done: False
step#: 19 reward: -0.0 total_reward: 0.0 done: False
step#: 20 reward: -0.0 total_reward: 0.0 done: False
step#: 21 reward: -0.0 total_reward: 0.0 done: False
step#: 22 reward: -0.0 total_reward: 0.0 done: False
step#: 23 reward: -0.0 total_reward: 0.0 done: False
step#: 24 reward: -0.0 total_reward: 0.0 done: False
step#: 25 reward: -0.0 total_reward: 0.0 done: False
step#: 26 reward: 0.0231 total_reward: 0.0231 done: False
step#: 27 reward: 0.0624 total_reward: 0.0855 done: False
step#: 28 reward: 0.0559 total_reward: 0.1414 done: False
step#: 29 reward: 0.0511 total_reward: 0.1925 done: False
step#: 30 reward: 0.0467 total_reward: 0.2391 done: False
```

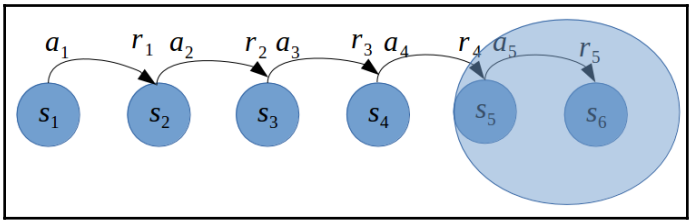
```
praveen@ubuntu: ~/rl_gym_book/ch7/carla-gym
(rl_gym_book) praveen@ubuntu:~/rl_gym_book/ch7/carla-gym$ python
Python 3.5.5 |Anaconda, Inc.| (default, Apr 26 2018, 13:47:34)
[GCC 7.2.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import gym
>>> import carla_gym
>>> env = gym.make("Carla-v0")
>>> obs = env.reset()
Initializing new Carla server...
Start pos 36 ([0.0, 3.0]), end 40 ([0.0, 3.0])
Starting new episode...
```

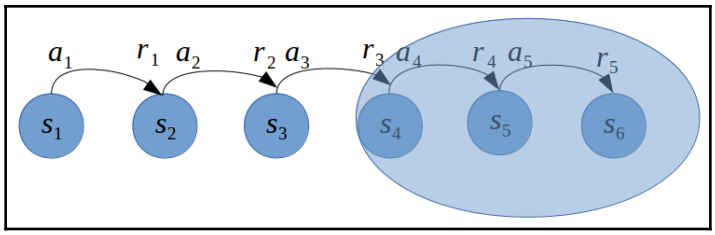
Chapter 8: Implementing an Intelligent - Autonomous Car Driving Agent using Deep Actor-Critic Algorithm

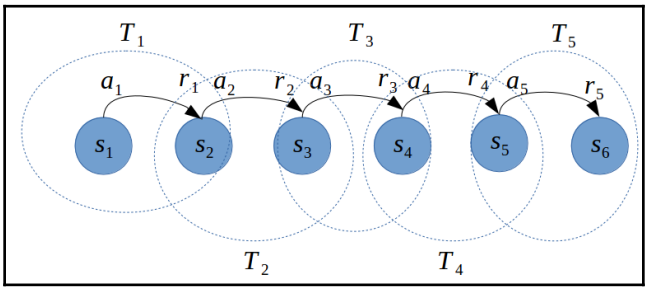


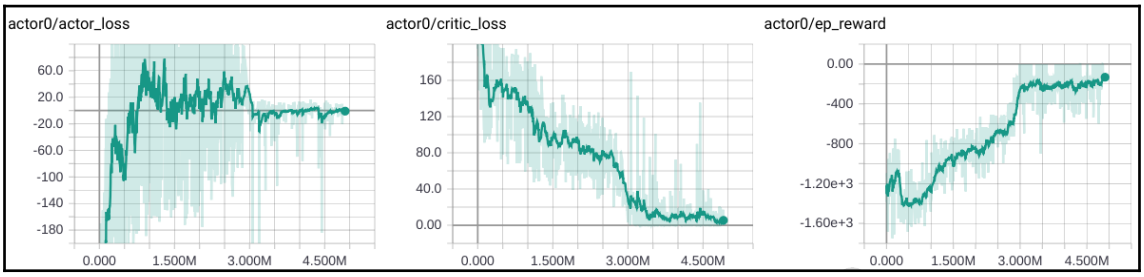


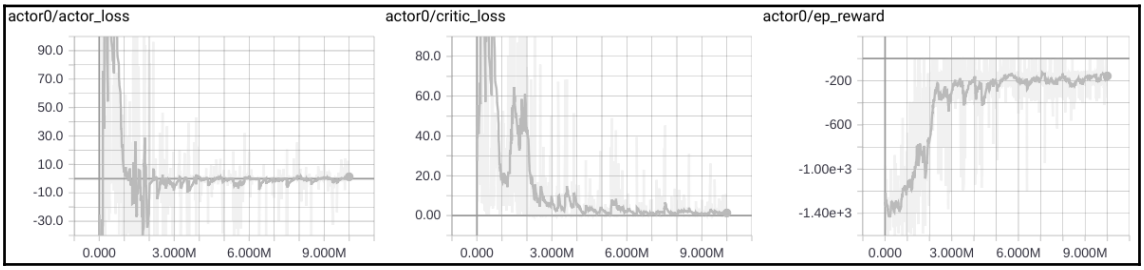


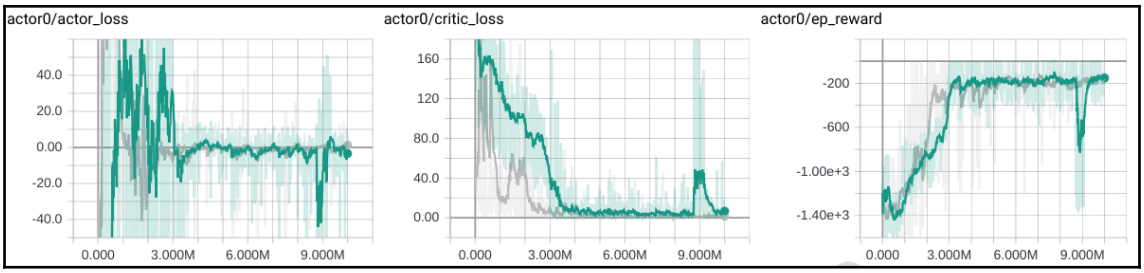



a_1 r_1 a_2 r_2 a_3 r_3 a_4 r_4 a_5 r_5
s_1 s_2 s_3 s_4 s_5 s_6

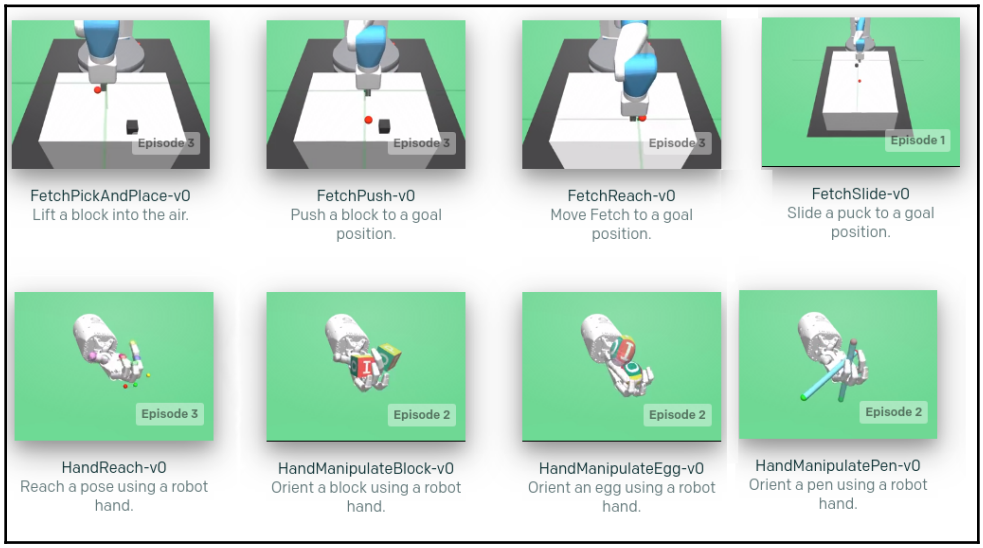

T_1 T_3 T_5
a_1 r_1 a_2 r_2 a_3 r_3 a_4 r_4 a_5 r_5
s_1 s_2 s_3 s_4 s_5 s_6
T_2 T_4

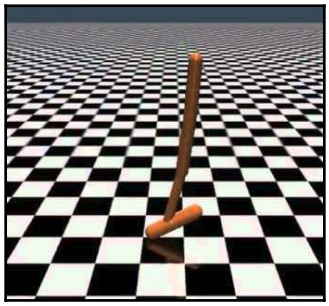

actor0/actor_loss
60.0 20.0 -20.0 -60.0 -100 -140 -180
0.000 1.500M 3.000M 4.500M
actor0/critic_loss
160 120 80.0 40.0 0.00
0.000 1.500M 3.000M 4.500M
actor0/ep_reward
0.00 -400 -800 -1.20e+3 -1.60e+3
0.000 1.500M 3.000M 4.500M

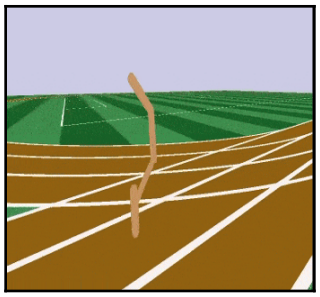

actor0/actor_loss
90.0 70.0 50.0 30.0 10.0 -10.0 -30.0
0.000 3.000M 6.000M 9.000M
actor0/critic_loss
80.0 60.0 40.0 20.0 0.00
0.000 3.000M 6.000M 9.000M
actor0/ep_reward
-200 -600 -1.00e+3 -1.40e+3
0.000 3.000M 6.000M 9.000M

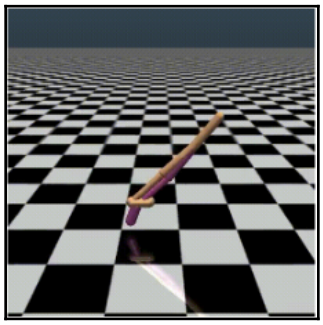

actor0/actor_loss
40.0
20.0
0.00
-20.0
-40.0
0.000
3.000M
6.000M
9.000M
actor0/critic_loss
160
120
80.0
40.0
0.00
0.000
3.000M
6.000M
9.000M
actor0/ep_reward
-200
-600
-1.00e+3
-1.40e+3
0.000
3.000M
6.000M
9.000M

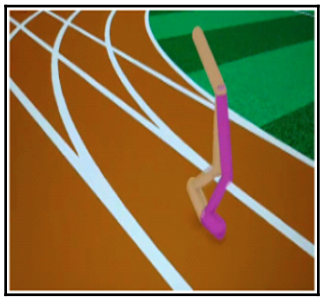
Chapter 9: Exploring the Learning Environment Landscape - Roboschool, Gym-Retro, StarCraft-II, DeepMindLab

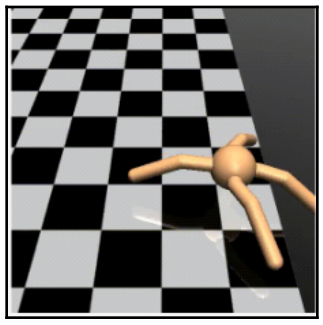


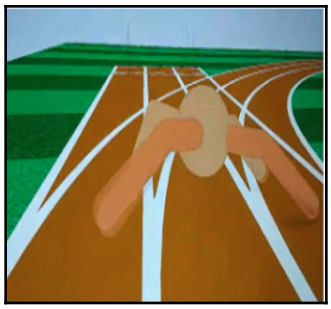


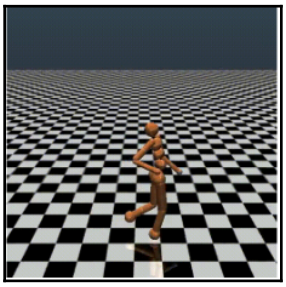


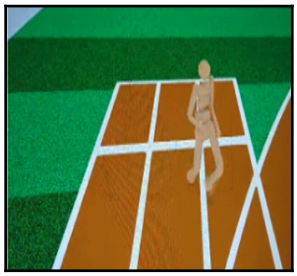


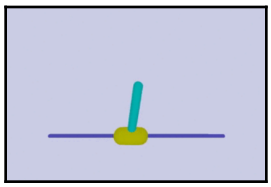


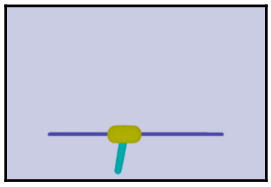


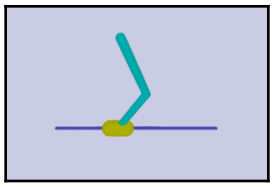


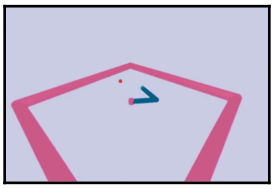


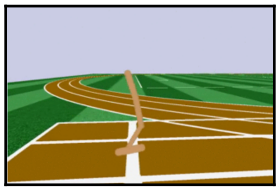


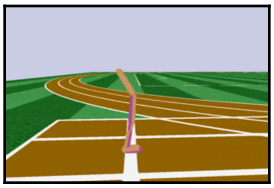


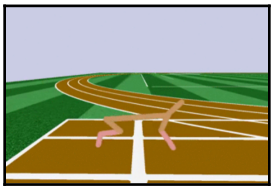


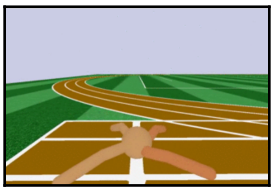


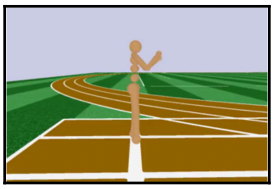


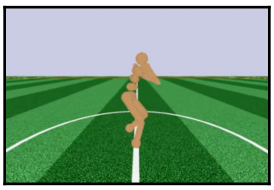


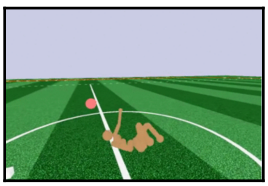


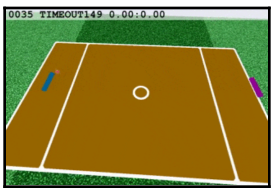


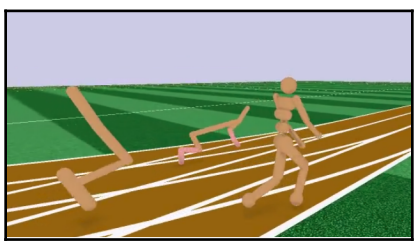




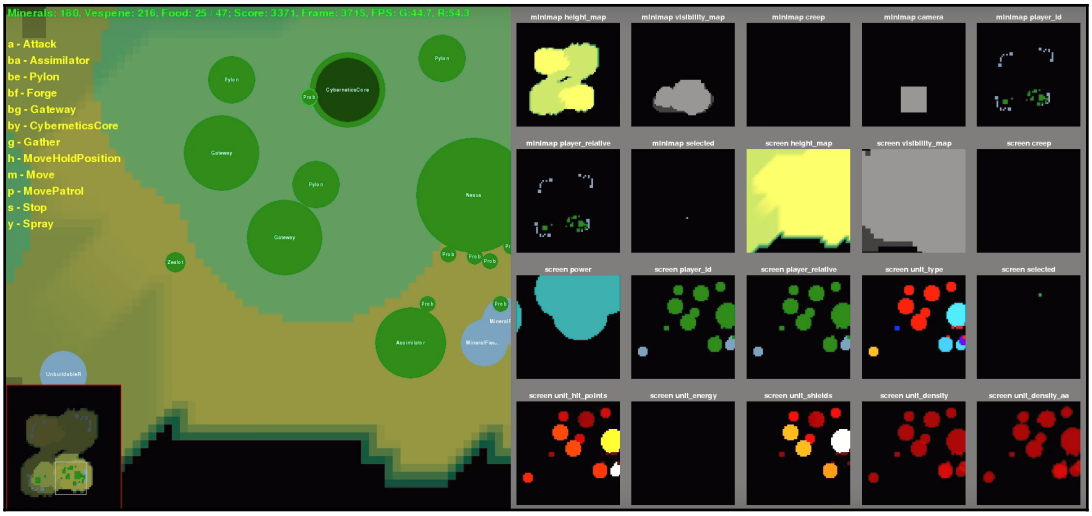




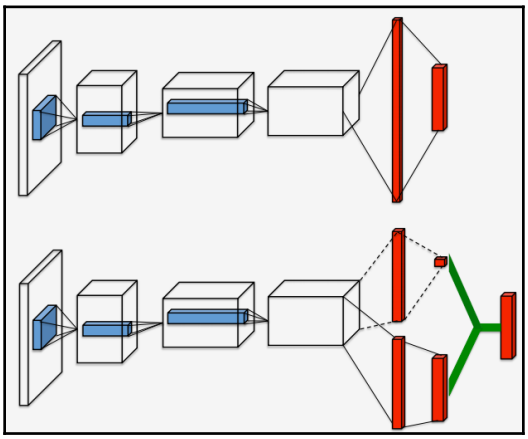

0035 TIMEOUT149 0.00:0.00

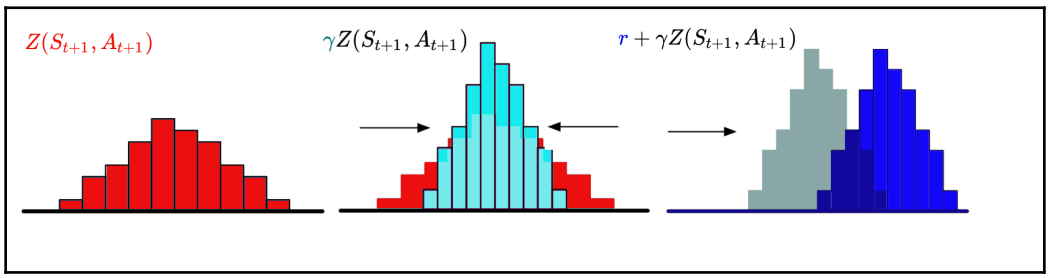



Lives: 3
Score: 40
Hi: 40


Minerals: 160, Vespene: 216, Food: 25 / 47, Score: 3371, Frame: 3715, FPS: G:44.7, R:54.3
a - Attack
ba - Assimilator
be - Pylon
bf - Forge
bg - Gateway
by - CyberneticsCore
g - Gather
h - MoveHoldPosition
m - Move
p - MovePatrol
s - Stop
y - Spray
minimap height_map
minimap visibility_map
minimap creep
minimap camera
minimap player_id
minimap player_relative
minimap selected
screen height_map
screen visibility_map
screen creep
screen power
screen player_id
screen player_relative
screen unit_type
screen selected
screen unit_hit_points
screen unit_energy
screen unit_shields
screen unit_density
screen unit_density_aa

Chapter 10: Exploring the Learning Algorithm Landscape - DDPG (Actor-Critic), PPO (Policy-Gradient), Rainbow (Value-Based)





Algorithm 1 Categorical Algorithm

input A transition $x_t, a_t, r_t, x_{t+1}, \gamma_t \in [0, 1]$

$Q(x_{t+1}, a) := \sum_i z_i p_i(x_{t+1}, a)$

$a^* \leftarrow \arg\max_a Q(x_{t+1}, a)$

$m_i = 0, \quad i \in 0, \dots, N-1$

for $j \in 0, \dots, N-1$ **do**

Compute the projection of $\hat{\mathcal{T}} z_j$ onto the support $\{z_i\}$

$\hat{\mathcal{T}} z_j \leftarrow [r_t + \gamma_t z_j]_{V_{\text{MIN}}}^{V_{\text{MAX}}}$

$b_j \leftarrow (\hat{\mathcal{T}} z_j - V_{\text{MIN}})/\Delta z \quad \# \, b_j \in [0, N-1]$

$l \leftarrow \lfloor b_j \rfloor, u \leftarrow \lceil b_j \rceil$

Distribute probability of $\hat{\mathcal{T}} z_j$

$m_l \leftarrow m_l + p_j(x_{t+1}, a^*)(u - b_j)$

$m_u \leftarrow m_u + p_j(x_{t+1}, a^*)(b_j - l)$

end for

output $-\sum_i m_i \log p_i(x_t, a_t)$ # Cross-entropy loss