



- ✓ 1. Introduction
- ✓ 2. The Setting, Revisited
- ✓ 3. Episodic vs. Continuing Tasks
- ✓ 4. Quiz: Test Your Intuition
- ✓ 5. Quiz: Episodic or Continuing?
- ✓ 6. The Reward Hypothesis
- ✓ 7. Goals and Rewards, Part 1
- ✓ 8. Goals and Rewards, Part 2
- ✓ 9. Quiz: Goals and Rewards
- ✓ 10. Cumulative Reward
- ✓ 11. Discounted Return
- ✓ 12. Quiz: Pole-Balancing
- ✓ 13. MDPs, Part 1
- ✓ 14. MDPs, Part 2
- ✓ 15. Quiz: One-Step Dynamics, Par...
- ✓ 16. Quiz: One-Step Dynamics, Par...
- ✓ 17. MDPs, Part 3
- ✓ 18. Finite MDPs
- 19. Summary

Finite MDPs

Please use [this link](#) to peruse the available environments in OpenAI Gym.

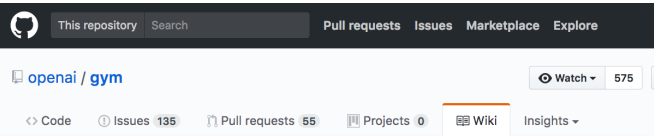


Table of environments

Marco Birk edited this page 6 days ago · 6 revisions

Here is a synopsis of the environments as of 2016-06-20, in order by space dimensionality. See discussion and code in [Write more documentation about environments: Issue #106](#).

Environment Id	Observation Space	Action Space	Rewa Rang
TimePilot-ram-v0	Box(128,)	Discrete(10)	(-inf, inf)
Amidar-ram-v0	Box(128,)	Discrete(10)	(-inf, inf)

The environments are indexed by **Environment Id**, and each environment has corresponding **Observation Space**, **Action Space**, **Reward Range**, **tStepL**, **Trials**, and **rThresh**.

CartPole-v0

Find the line in the table that corresponds to the **CartPole-v0** environment. Take note of the corresponding **Observation Space** (**Box(4,)**) and **Action Space** (**Discrete(2)**).

CartPole-v0	Box(4,)	Discrete(2)	(-inf, inf)
-----------------------------	---------	-------------	-------------

As described in the [OpenAI Gym documentation](#),



valid actions and observations.

- The **Discrete** space allows a fixed range of non-negative numbers.
- The **Box** space represents an n-dimensional box, so valid actions or observations will be an array of n numbers.

Observation Space

The observation space for the CartPole-v0 environment has type **Box(4,)**. Thus, the observation (or state) at each time point is an array of 4 numbers. You can look up what each of these numbers represents in [this document](#). After opening the page, scroll down to the description of the observation space.

Observation

Type: Box(4)

Num	Observation	Min	Max
0	Cart Position	-2.4	2.4
1	Cart Velocity	-Inf	Inf
2	Pole Angle	~ -41.8°	~ 41.8°
3	Pole Velocity At Tip	-Inf	Inf

Notice the minimum (-Inf) and maximum (Inf) values for both **Cart Velocity** and the **Pole Velocity at Tip**.

Since the entry in the array corresponding to each of these indices can be any real



Action Space

The action space for the CartPole-v0 environment has type **Discrete(2)**. Thus, at any time point, there are only two actions available to the agent. You can look up what each of these numbers represents in [this document](#) (note that it is the same document you used to look up the observation space!). After opening the page, scroll down to the description of the action space.

Actions

Type: Discrete(2)

Num	Action
0	Push cart to the left
1	Push cart to the right

In this case, the action space \mathcal{A} is a finite set containing only two elements.

Finite MDPs

Recall from the previous concept that in a finite MDP, the state space \mathcal{S} (or \mathcal{S}^+ , in the case of an episodic task) and action space \mathcal{A} must both be finite.

Thus, while the CartPole-v0 environment does specify an MDP, it does not specify a **finite** MDP. In this course, we will first learn how to solve finite MDPs. Then, later in this course, you will learn how to use neural networks to solve much more complex

