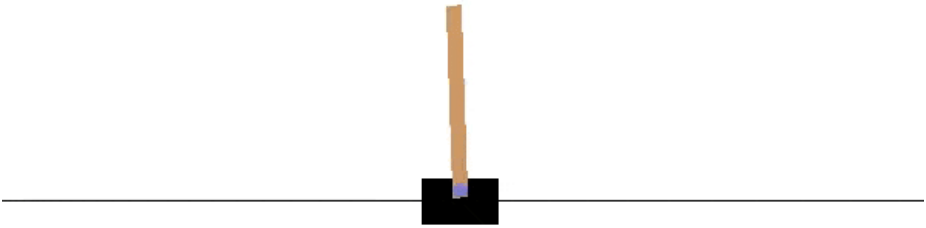
Quiz: Pole-Balancing



Source: https://medium.com/@tuzzer/cart-pole-balancing-with-q-learning-b54c6068d947

In this classic reinforcement learning task, a cart is positioned on a frictionless track, and a pole is attached to the top of the cart. The objective is to keep the pole from falling over by moving the cart either left or right, and without falling off the track.

In the OpenAI Gym implementation, the agent applies a force of +1 or -1 to the cart at every time step. It is formulated as an episodic task, where the episode ends when (1) the pole falls more than 20.9 degrees from vertical, (2) the cart moves more than 2.4 units from the center of the track, or (3) when more than 200 time steps have elapsed. The agent receives a reward of +1 for every time step, including the final step of the episode. You can read more about this environment in OpenAI's github. This task also appears in Example 3.4 of the textbook.

Quiz Question

Recall that the agent receives a reward of +1 for every time step, including the final step of the episode. Which discount rates would encourage the agent to keep the pole balanced for as long as possible? (Select all that apply.)

- The discount rate is 1.
- The discount rate is 0.9.

- The discount rate is 0.5.

Submit

Quiz Question

Say that the reward signal is amended to only give reward to the agent at the end of an episode. So, the reward is 0 for every time step, with the exception of the final time step. When the episode terminates, the agent receives a reward of **-1**. Which discount rates would encourage the agent to keep the pole balanced for as long as possible? (Select all that apply.)

- ☐

  The discount rate is 1.

  ✓

- The discount rate is 0.9.
- The discount rate is 0.5.
- ☐

  (None of these discount rates would help the agent, and there is a problem with the reward signal.)

  ✓

Submit

Quiz Question

Say that the reward signal is amended to only give reward to the agent at the end of an episode. So, the reward is 0 for every time step, with the exception of the final time step. When the episode terminates, the agent receives a reward of **+1**. Which discount rates would encourage the agent to keep the pole balanced for as long as possible? (Select all that apply.)

- ☐

  The discount rate is 1.

  ✓

- ☐

  The discount rate is 0.9.

  ✓
- ☐

  The discount rate is 0.5.

  ✓
- (None of these discount rates would help the agent, and there is a problem with the reward signal.)

Submit