Pendulum-v0

1. There are three observation inputs for this environment, representing the angle of the pendulum cos(θ) [-1.0, 1.0], sin(θ)[-1.0,1.0], and its angular velocity dθ[-8.0, 8.0].
2. The action is a value between -2.0 and 2.0, representing the amount of left or right force on the pendulum.
3. The precise equation for reward: `-(theta^2 + 0.1*theta_dt^2 + 0.001*action^2)`
4. The reset method generates random angle from -pi to pi, and random velocity between -1 and 1
5. There is no specified termination.

MountainCar-v0

1. There are two observation inputs for this environment, representing the positioin of the car [-1.2, 0.6], and velocity [-0.07, 0.07].

2. The action is a descreet value: 0 — push left, 1 — no push, 2 — push right.

3. Reward of 0 is awarded if the agent reached the flag (position = 0.5) on top of the mountain. Reward of -1 is awarded if the position of the agent is less than 0.5.

4.The reset method generates random position of car [-0.6,-0.4] with zero velocity

5. Termination: the car position is more than 0.5 and episode length is greater than 200