LunarLander-v2

1. Observation vector (8 dimensions) consists of lander’s coordinates (horizontal, vertical), speeds (horizontal, vertical), angle, angular speed, contact of the first leg: 1 – there is contact, 0- no contact, contact of the second leg.
2. Action vector takes values from 0 to 3. 0 – do nothing, 1 - fire left orientation engine, 2 - fire main engine, 3 - fire right orientation engine.
3. Reward is calculated depending on the performance of the lander. If lander moves away from landing pad it loses the reward. The lander gets -100 (+100) if its fails (manages) to land between the flags.
4. Each reset gives the lander a random position.
5. Termination happens when the landing has happened and the lander receives additional -100 or +100 points depending on the success of completing the task.

Acrobot-v1

1. Observation vector consists of the sin() and cos() of two rotational joint angles and joint angular velocities: [cos(theta1) sin(theta1) cos(theta2) sin(theta2) thetaDot1 thetaDot2].
2. Action vector takes values from 0 to 2: the value is the applied torque on the joint between two pendulum links.
3. No specific formula for the reward. Takes -1 value if the acrobot fails to reach the goal.
4. Each reset randomizes the pendulum configuration.
5. Termination doesn’t happen until the end of the lower link goes up to a given height. In case if acrobot fails to reach the upper position, the episode stops automatically after 499 steps.