The state-space S is continuous and consists of three states [cos(theta),cos(theta),theta-point], where theta= 0 when the pendulum is oriented upwards. Action-space A is also continuous. There is only one action, namely the joint effort. By using the action, we can change the state of the agent.

Our goal is to direct the pendulum upwards to where the theta= 0. To achieve this, we try to minimize the costs:

(theta^2 + 0.1\*theta\_dt^2 + 0.001\*action^2).

The first term (theta^2) aims to make theta as small as possible. (0.1\*theta\_dt^2) aims to reduce the speed and plays a role like a damper. The last term finally serves to reduce the joint effort required to reach the goal.