

환경. = $\left[\text{현재 State, 다음 State, Action, Action 결과, reward} \dots \right]$
 기억함.

Optimizer = Adam (손실 감소속도가 가장 빠름)

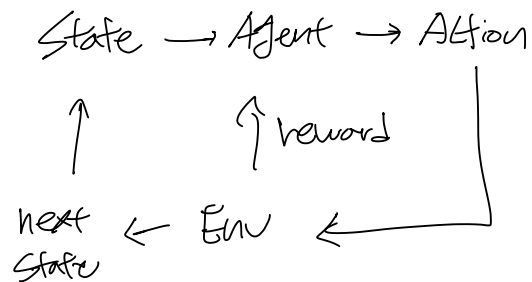
Activation Function = ReLU

Gamma = 0.009

buffer_size = 5000

learning_rate = 0.0001

batch_size = 64



buffer를 사용하여 $S_t, S_{t+1}, \text{Action}, \text{reward}, \text{Episode 종료 여부}$ 저장.

→ Sample 하기

