DQN

1. 初始化经验缓冲区，容量为
2. 初始化策略动作值函数，即带有初始权重的𝑄网络
3. 初始化目标动作值函数，即带有初始权重的网络，且
4. 执行个回合，对于每个回合
5. 初始化环境，得到初始状态
6. 循环个时间步长，对于每个时步
7. 使用−贪心策略选择动作
8. 环境根据反馈奖励和下一个状态
9. 更新状态
10. 存储转移即到经验回放中
11. 更新策略如下：
12. 1) 从中随机采样一个小批量的转移
13. 2) 计算实际的值
14. 3) 对损失函数关于参数做随机梯度下降
15. 4) 每步重置