# DQN算法

正式进入使用神经网络的阶段。由于Q_learning的存表方式只适合离散且空间小的情况，因此DQN提出使用神经网络去拟合Q值。

**DQN只能应用在离散的动作空间中，因为他要取max**

DQN使用Q_learning的时序差分向构建损失函数：

那么 Q 网络的损失函数是什么呢？我们先来回顾一下 Q-learning 的更新规则（参见 5.5 节）：

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[ r + \gamma \max_{a' \in A} Q(s',a') - Q(s,a) \right]$$

上述公式用**时序差分**（temporal difference，TD）学习目标$r + \gamma \max_{a' \in A} Q(s',a')$来增量式更新$Q(s,a)$，也就是说要使$Q(s,a)$和 TD 目标$r + \gamma \max_{a' \in A} Q(s',a')$靠近。于是，对于一组数据$\{(s_i, a_i, r_i, s_i')\}$，我们可以很自然地将 Q 网络的损失函数构造为均方误差的形式：

$$\omega^* = \arg\min_{\omega} \frac{1}{2N} \sum_{i=1}^{N} \left[ Q_\omega(s_i, a_i) - \left( r_i + \gamma \max_{a'} Q_\omega(s_i', a') \right) \right]^2$$