动态规划寻找最优策略

本讲着重讲解利用动态规划来进行强化学习，具体是进行强化学习中的“规划”，也就是在已知模型的基础上判断一个策略的价值函数，并在此基础上寻找到最优的策略和最优价值函数，或者直接寻找最优策略和最优价值函数。本讲是整个强化学习课程核心内容的引子。

**简介**

动态规划算法是解决复杂问题的一个方法，算法通过把复杂问题分解为子问题，通过求解子问题进而得到整个问题的解。在求解子问题时，它的结果通常需要存储起来被用来解决后续复杂问题。

当问题具有下列特性时，通常可以考虑使用动态规划来求解：

* 一个复杂问题的最优解由数个小问题的最优解构成，可以通过寻找子问题的最优解来得到复杂问题的最优解。
* 子问题在复杂问题内重复出现，使得子问题的解可以被存储起来重复利用。

马尔可夫决定过程(MDP)具有上述两个属性：Bellman方程把问题递归为求解子问题，价值函数就相当于存储了一些子问题的解，可以复用。因此，可以使用动态规划算法来求解MDP。