强化学习 寻找 Blackjack 不记牌情况下的 Best policy

（1）规则介绍

使用除大小王之外的 52 张牌，游戏者的目标是使手中的牌的点数之和不超过 21 点且尽量大，超过 21 点称 为“爆牌”。

牌的点数划分如下：

数字牌：2 到 10 的牌按面值计算点数；

面牌（J、Q、K）：每张值 10 点；

A（Ace）：可以算作 1 点或 11 点，取决于哪种计算方式对玩家更有利。

假设玩家与庄家独立竞争。

游戏流程

游戏开始时，每人发两张牌。玩家的两张牌面朝上，而庄家则有一张面朝上（明牌）和一张面朝下 （暗牌）。 如果玩家立即拿到 21 点（一张 Ace 和一张 10 点），则称为天和。若庄家也有天和，则平局，否则玩 家直接获胜。 如果玩家没有天和，那么他可以一张一张地要额外的牌，直到他停牌或超过 21 点。如果玩家“爆牌”，则游戏失败；如果玩家不要牌，则轮到庄家。 庄家根据固定策略要牌或不要牌：点数总和大于或等于 17 点时停牌，否则继续要牌。 如果庄家“爆牌”，则玩家获胜；否则，点数总和更接近 21 点一方的获胜，相同则平局。

状态空间

玩家手牌点数总和 (12 - 21) ;

庄家明牌点数 (ace - 10) ;

是否有 Useable Ace (Yes or No) ;

动作空间

要牌（twist）：玩家要多一张牌。

停牌（stick）：玩家不要求额外的牌， 之后不再做出行动。

（2）要求

实现以下四种强化学习方法用以解决这个问题，并对比四种方法的优劣 Monte Carlo Sarsa Q-Learning Deep Q-Learning 计算并绘制最优 policy 下的 value function.