假期计划：

1. 了解机器学习，深度学习，强化学习及深度强化学习的概念及背景。
2. 搞懂强化学习关键要素：agent（智能体），reward（奖励），action（行为），state（状态），environment（环境）。

* agent：主要涉及到：策略（Policy），价值函数（Value Function）和模型（Model）。Policy，可以理解为行动指南，让agent执行什么动作，在数学上可以理解为从状态state到动作action的映射，可分为确定性策略（Deterministic policy）和随机性策略（Stochastic policy），前者是指在某特定状态下执行某个特定动作，后者是根据概率来执行某个动作。Value Function，对未来总Reward的一个预测。Model，一个对环境的认知框架，可以预测采取动作后的下一个状态是什么，很多情况下是没有模型的，agent只能通过与环境互动来提升策略。
* state：可以细分为三种，Environment State，Agent State和Information State。Environment State是agent所处环境包含的信息，简单理解就是很多特征数据，也包含了无用的数据。Agent State是输入给agent的信息，也就是特征数据。Information State是一个概念，即当前状态包含了对未来预测所需要的有用信息，过去信息对未来预测不重要，该状态就满足马尔科夫性（Markov Property）。Environment State，Agent State都可以是Markov Property。
* environment：可以分为完全可观测环境（Fully Observable Environment）和部分可观测环境（Partially Observable Environment）。Fully Observable Environment就是agent了解了整个环境，显然是一个理想情况。Partially Observable Environment是agent了解部分环境的情况，剩下的需要靠agent去探索。

1. 了解强化学习的一些典型案例，如TD-gammon。
2. 学习强化学习的一些算法，强化学习算法按照agent分类，可以分为下面几类：

* 关注最优策略（Policy based）
* 关注最优奖励总和（Value based）
* 关注每一步的最优行动（Action based）

5.了解DQN模型，它用到了两个关键技术：

* 样本池（Experience Reply）：将采集到的样本先放入样本池，然后从样本池中随机选出一条样本用于对网络的训练。这种处理打破了样本间的关联，使样本间相互独立。
* 固定目标值网络（Fixed Q-target）：计算网络目标值需用到现有的Q值，现用一个更新较慢的网络专门提供此Q值。这提高了训练的稳定性和收敛性。