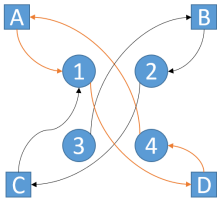
知识点总结（博弈）

- 强化学习、监督学习和深度卷积神经网络学习的描述：评估学习方式、有标注信息学习方式、端到端学习方式
- 在状态s，按照某个策略行动后在未来所获得回报值的期望”，这句话描述了状态s的价值函数
- 在状态s，按照某个策略采取动作a后在未来所获得回报值的期望”，这句话描述了状态s的动作-价值函数
- 博弈相关概念
  - 博弈的要素
    - 玩家 (player)： 参与博弈的决策主体
    - 策略 (strategy)： 玩家可以采取的行动方案，是一整套在采取行动之前就已经准备好的完整方案。
      - 某个玩家可采纳策略的全体组合形成了策略集 (strategy set)
      - 所有玩家各自采取行动后形成的状态被称为局势 (outcome)
      - 混合策略 (mixedstrategy): 玩家可通过一定概率来选择若干个不同的策略
      - 纯策略 (pure strategy): 玩家每次行动都选择某个确定的策略
    - 收益 (payoff)： 各个玩家在不同局势下得到的利益
      - 混合策略意义下的收益应为期望收益 (expected payoff)
    - 规则 (rule)： 对玩家行动的先后顺序、玩家获得信息多少等内容的规定
  - 在囚徒困境中，最优解为两人同时沉默　但是两人实际倾向于选择同时认罪 (均衡解)
  - 博弈的分类
    - 合作(cooperative)博弈与非合作(non-cooperative)博弈
      - 合作：部分玩家可以组成联盟以获得更大的收益
      - 非合作：玩家在决策中都彼此独立，不事先达成合作意向
    - 静态(static)博弈与动态(dynamic)博弈
      - 静态：所有玩家同时决策，或玩家互相不知道对方的决策
      - 动态：玩家所采取行为的先后顺序由规则决定，且后行动者知道先行动者所采取的行为
    - 完全信息(complete information)博弈与不完全(incomplete)信息博弈
      - 完全信息：所有玩家均了解其他玩家的策略集、收益等信息
      - 不完全信息： 并非所有玩家均掌握了所有信息
  - 纳什均衡
    - 博弈的稳定局势即为纳什均衡 (Nash equilibrium)
    - 玩家所作出的这样一种策略组合: 任何玩家单独改变策略都不会得到好处。

- Nash定理：若玩家有限，每位玩家的策略集有限，收益函数为实值函数，则博弈必存在混合策略意义下的纳什均衡。
- 纳什均衡的本质：不后悔

- 遗憾最小化算法
  - 根据过去博弈中的遗憾程度来决定将来动作选择的方法
  - 玩家⊠在过去⊠轮中采取策略σ_i的累加遗憾值为：
  - $Regret_i^T(\sigma_i) = \sum_{t=1}^{T}(\mu_i(\sigma_i, \sigma_{-i}^t) - \mu_i(\sigma^t))$
  - 在第T+1轮次玩家i选择策略a的概率如下 (悔值越大越选择)
  - $P(a) = \frac{Regret_i^T(a)}{\sum_{b \in \Sigma_i} Regret_i^T(b)}$
  - 双边匹配算法
    - 需要双向选择的情况被称为是双边匹配问题，需要双方互相满足对方的需求才会达成匹配
    - 稳定匹配是指没有任何人能从偏离稳状态中获益
    - 针对双边稳定匹配问题的算法并应用于稳定婚姻问题的求解
  - 单边匹配算法
    - 一类交换不可分标的物的匹配问题
    - 最大交易圈算法 (TTC)：
      - **假设某寝室有A、B、C、D四位同学和1、2、3、4四个床位**
        - 当前给A、B、C、D四位同学随机分配4、3、2、1四个床位
        - 已知四位同学对床位偏好如下：

同学	偏好
A	1>2>3>4
B	2>1>4>3
C	1>2>4>3
D	4>3>1>2



        - 可以看出交易图中A和D之间构成一个交易圈
        -