

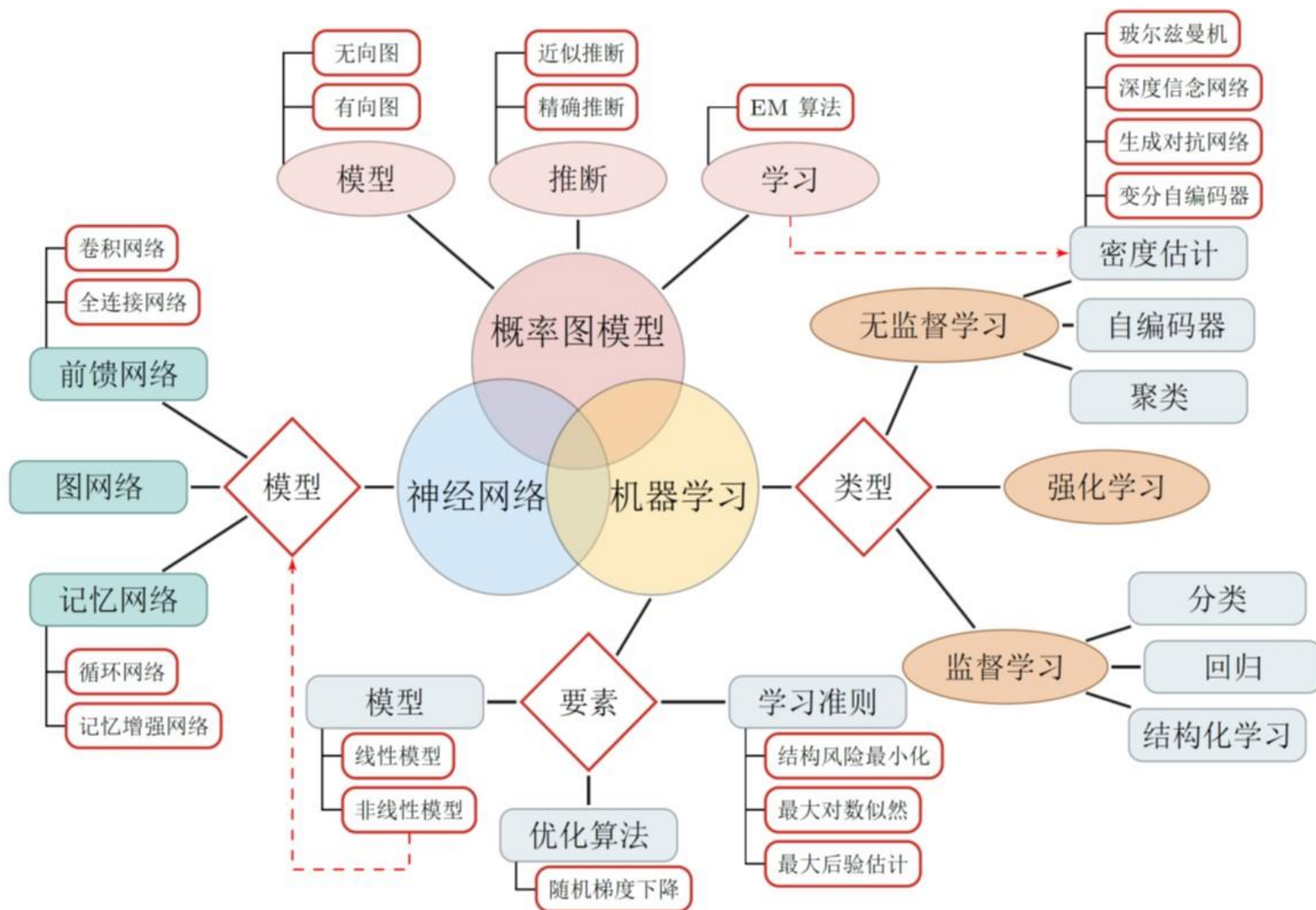


实验课内容总结

陈昱夫



学习模式





学习模式

	监督学习	无监督学习	强化学习
训练样本	训练集 $\{(\mathbf{x}^{(n)}, y^{(n)})\}_{n=1}^N$	训练集 $\{\mathbf{x}^n\}_{n=1}^N$	智能体和环境交互的 轨迹 τ 和累积奖励 G_τ
优化目标	$y = f(\mathbf{x})$ 或 $p(y \mathbf{x})$	$p(\mathbf{x})$ 或带隐变量 \mathbf{z} 的 $p(\mathbf{x} \mathbf{z})$	期望总回报 $\mathbb{E}_\tau[G_\tau]$
学习准则	期望风险最小化 最大似然估计	最大似然估计 最小重构错误	策略评估 策略改进



学习方法

传统机器学习

逻辑回归

K近邻

决策树

Adaboost算法

支持向量机

隐马尔科夫模型

贝叶斯方法

K-均值算法

深度学习

卷积神经网络 (CNN)

循环神经网络 (RNN)

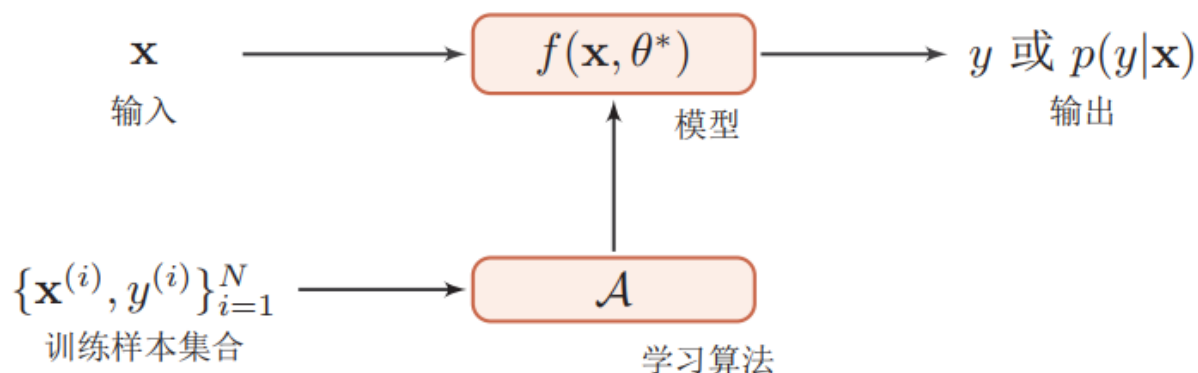
深度置信网络 (DBN)

受限玻尔兹曼机 (RBM)

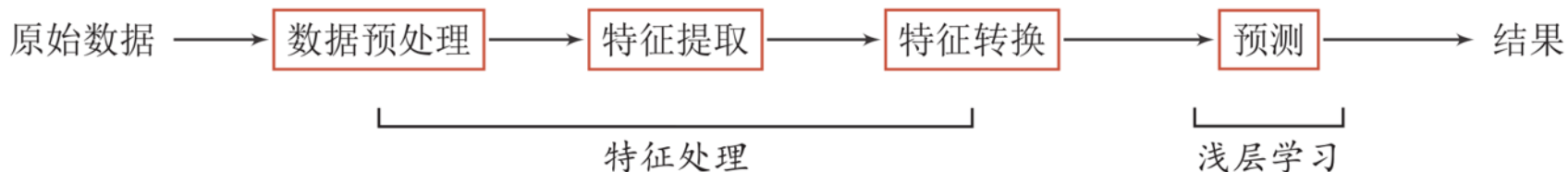


机器学习

- 传统的机器学习主要关注于如何学习一个预测模型。



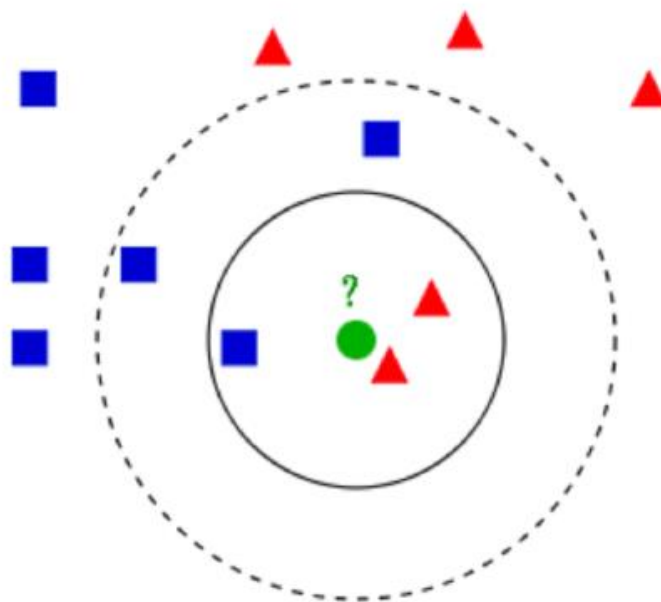
- 利用人类经验选取好的特征称之为特征工程 (Feature Engineering) 问题。





Lab1 K近邻算法

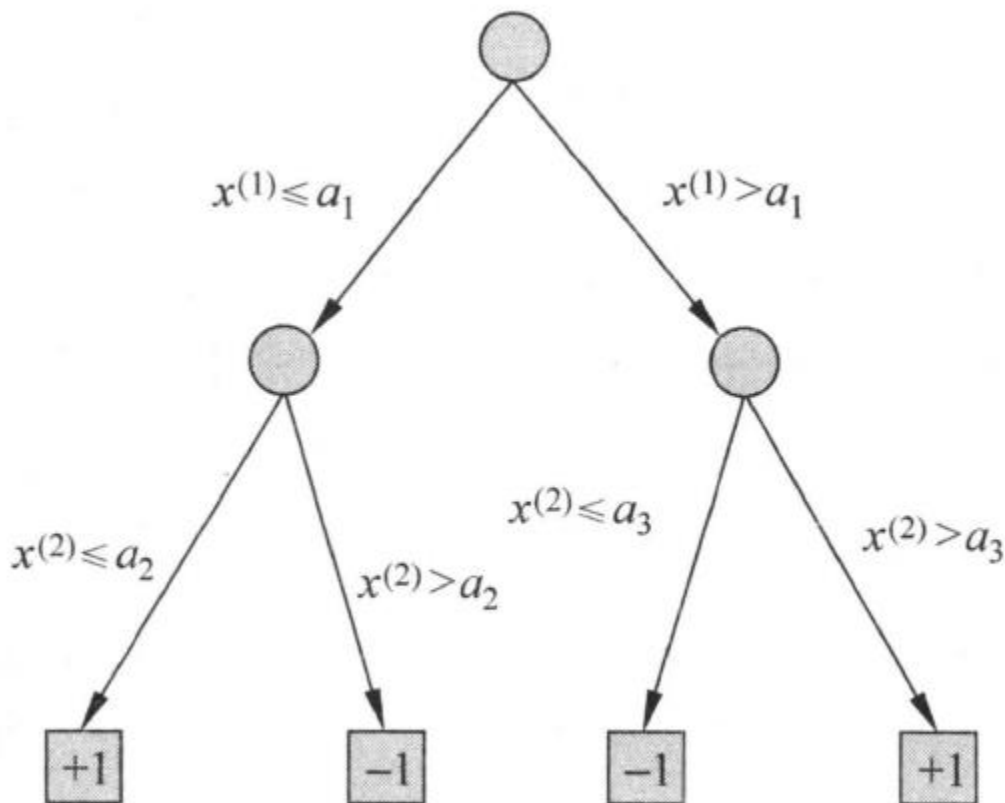
- 距离度量
 - 欧氏距离
 - 曼哈顿距离
- 分类决策规则
 - 多数投票
 - 加权求和





Lab2 决策树算法

- 特征选择
 - 信息增益: ID3
 - 信息增益率: C4.5
 - Gini指数: CART
- 决策树的修剪
 - 预剪枝
 - 后剪枝

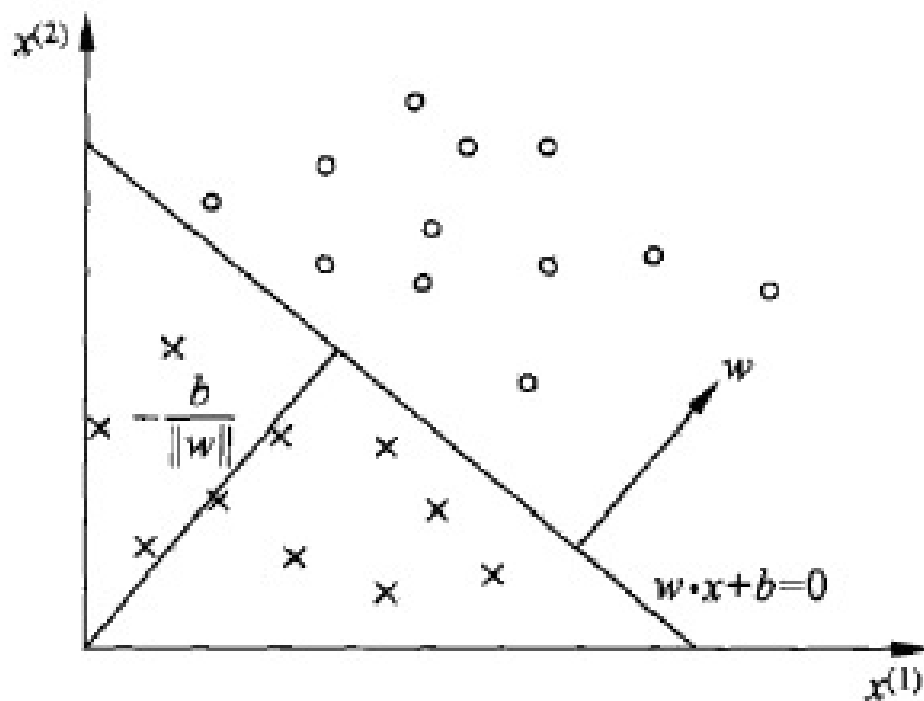




Lab3 感知机与逻辑回归

- 一种二类分类的线性分类模型。对应于特征空间中实例划分为正负两类的分离超平面，属于判别模型。

$$\text{sign} \left(\left(\sum_{i=1}^d w_i x_i \right) - \theta \right)$$

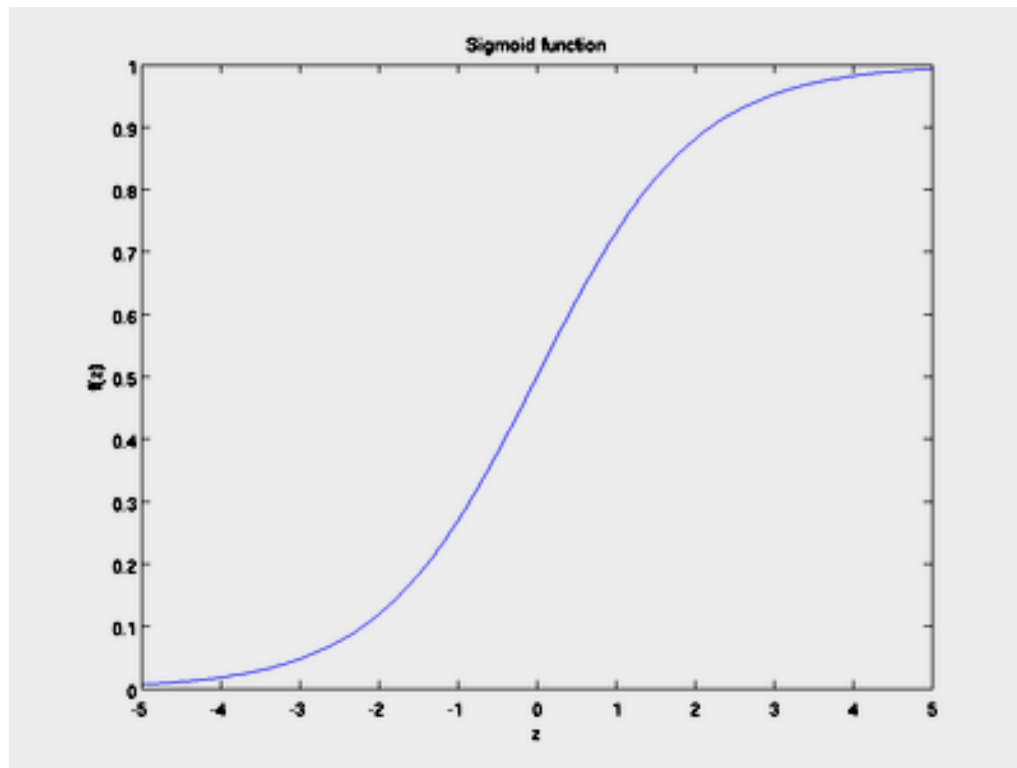




Lab3 感知机与逻辑回归

- 一种分类模型，由条件概率分布 $P(Y|X)$ 表示，形式为参数化的逻辑斯谛分布。

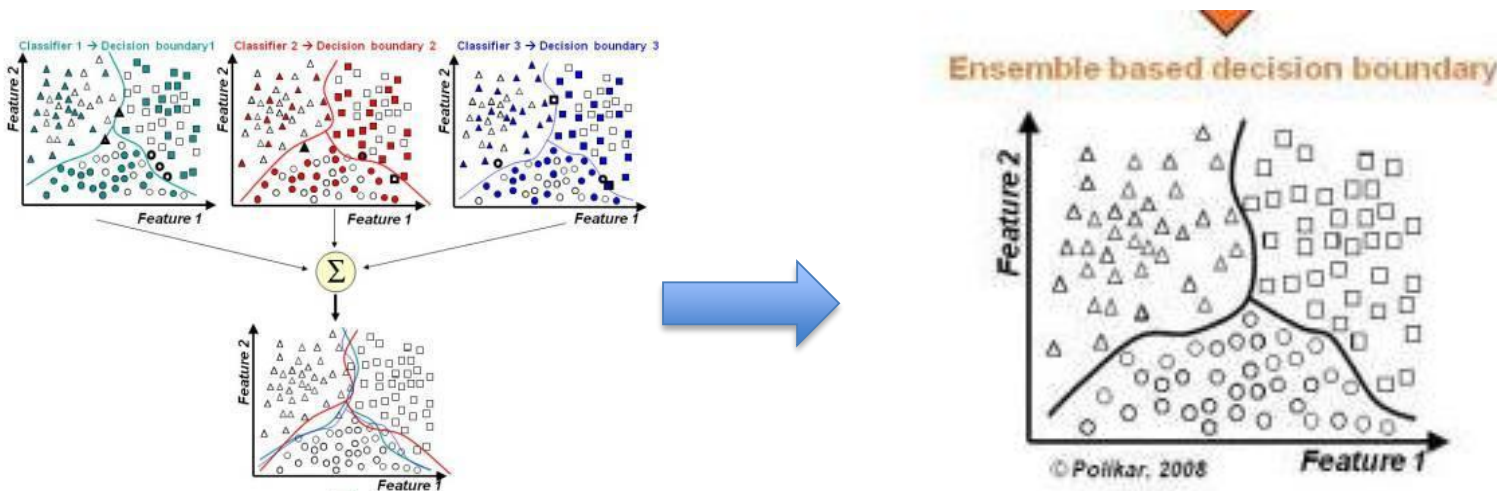
$$h(x) = \theta(x) = \frac{1}{1 + e^{-w^T x}}$$



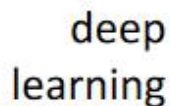


Lab8 集成学习

- 集成学习：通过某种策略将多个模型集成起来，群体决提高决策



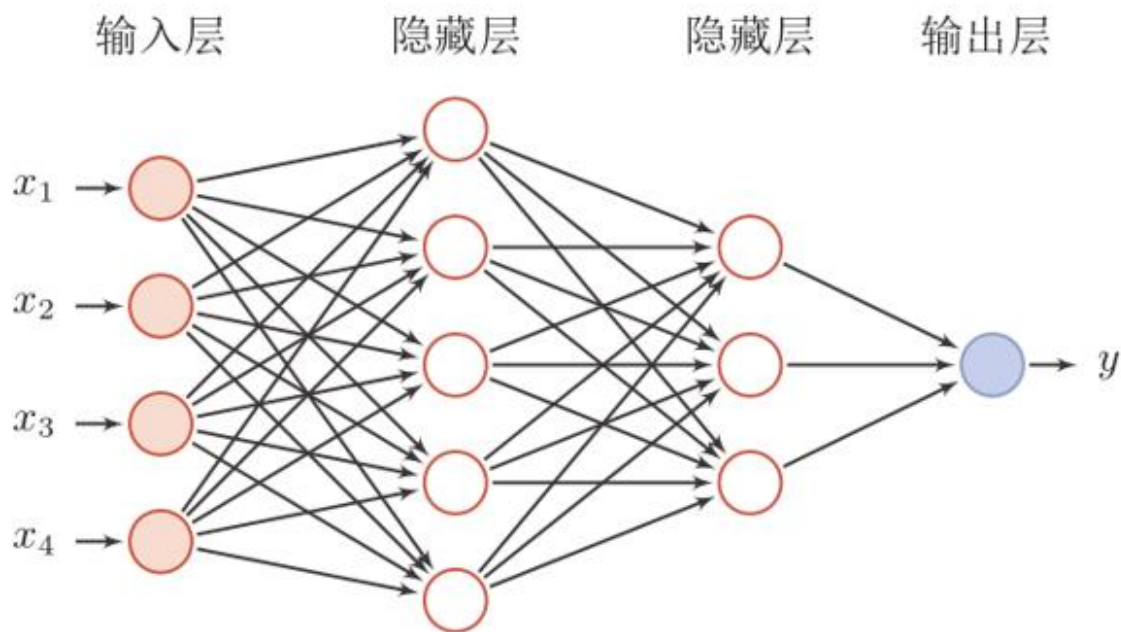
- 提升 (Boosting) 方法：通过改变训练样本的权重，学习多个分类器，并将这些分类器进行线性组合，提高分类的性能。





Lab4 简单神经网络

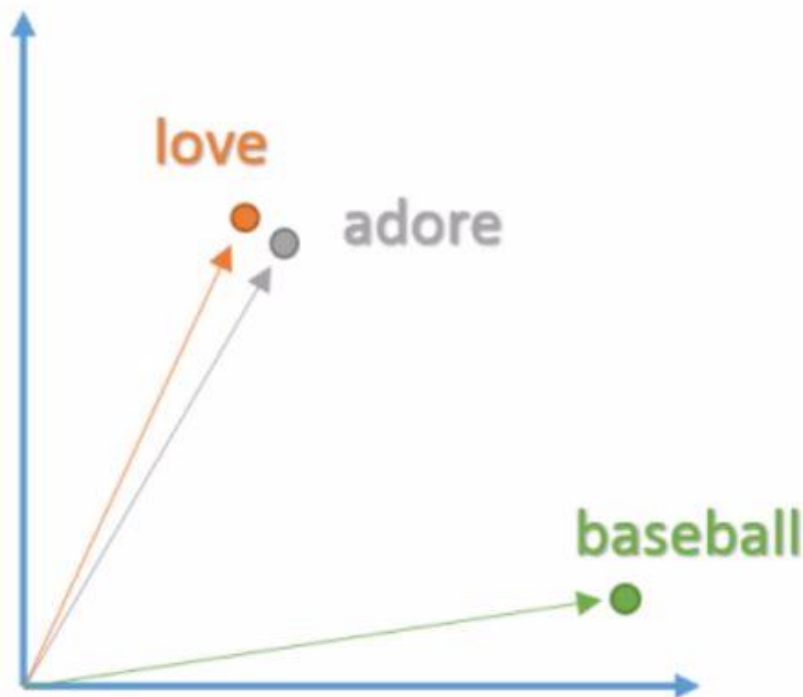
- 由神经元构成的多层结构。





Lab5 词嵌入

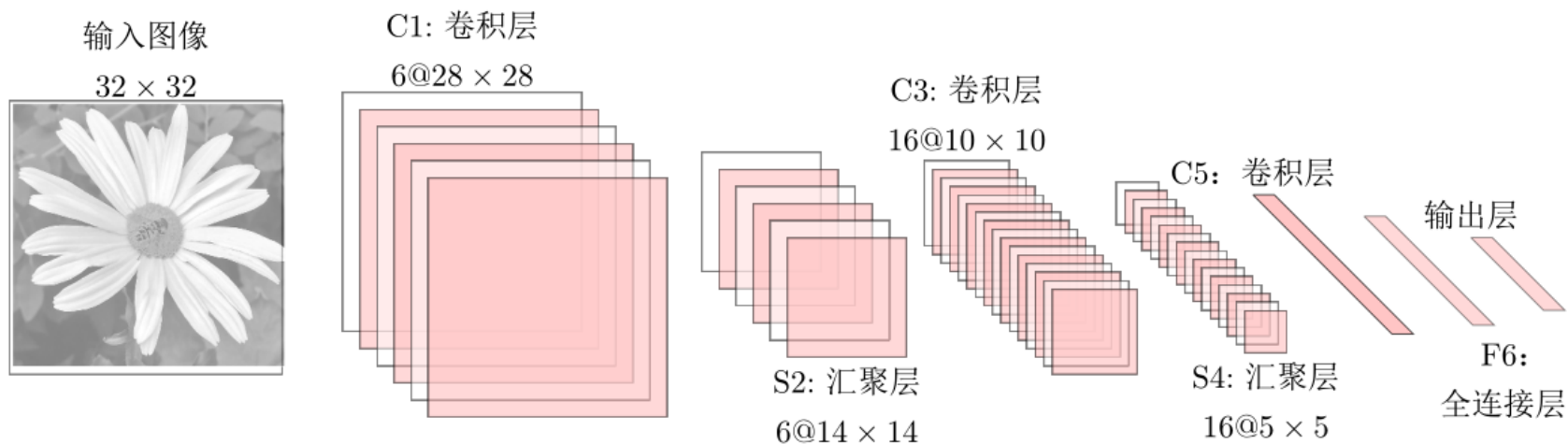
- 嵌入：将一个局部表示空间中的一些对象映射到另一个低维的分布式表示空间中，并尽可能保持不同对象之间的拓扑关系。
- 自然语言中词的分布式表示，也经常叫做词嵌入





Lab6 卷积神经网络

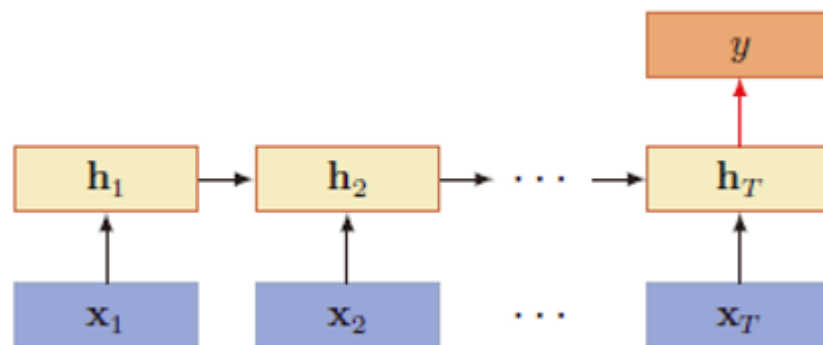
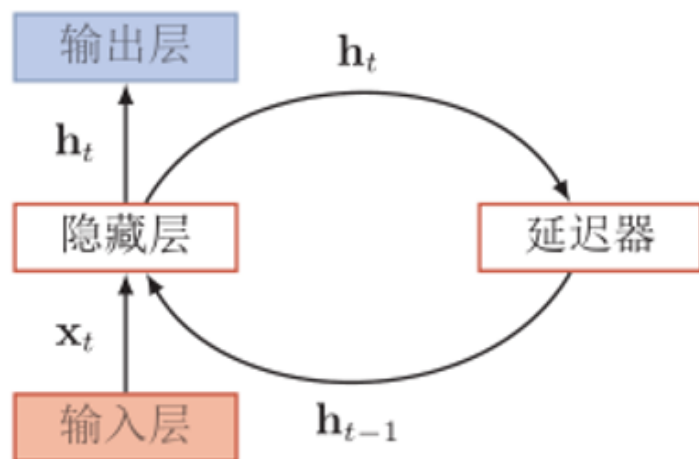
- 一种具有局部连接、权重共享等特性的深层前馈神经网络。





Lab7 递归神经网络

- 一类具有短期记忆能力的神经网络。

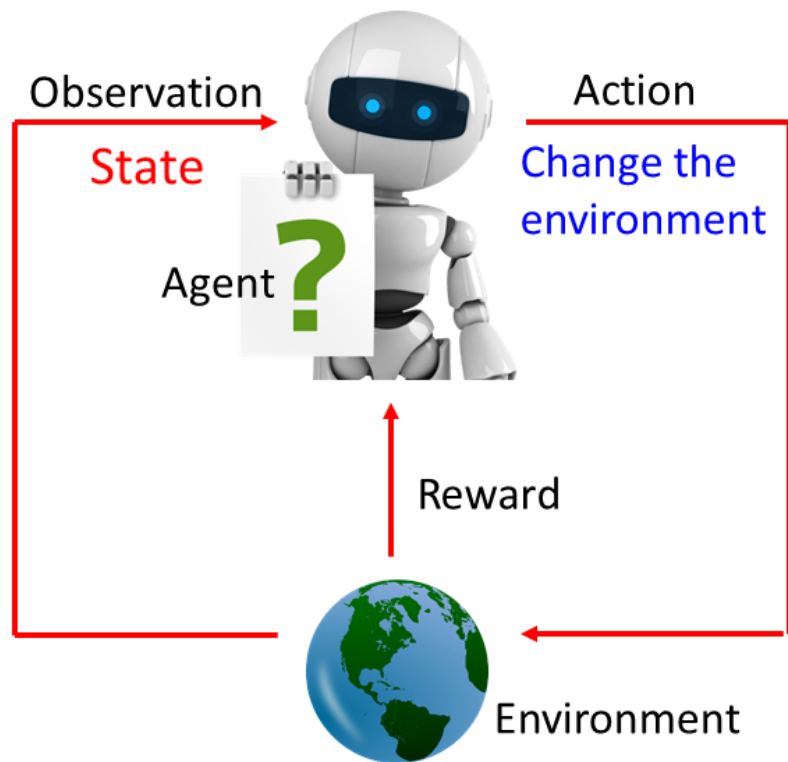




强化学习

- 一类从与环境的交互中不断学习的问题以及解决这类问题的方法
- 学习过程可以使用状态-行动链条来表示：

$$\{s_0, a_0, s_1, a_1, \dots, s_{t-1}, a_{t-1}, s_t\}$$





Lab9 盲目搜索与启发式搜索

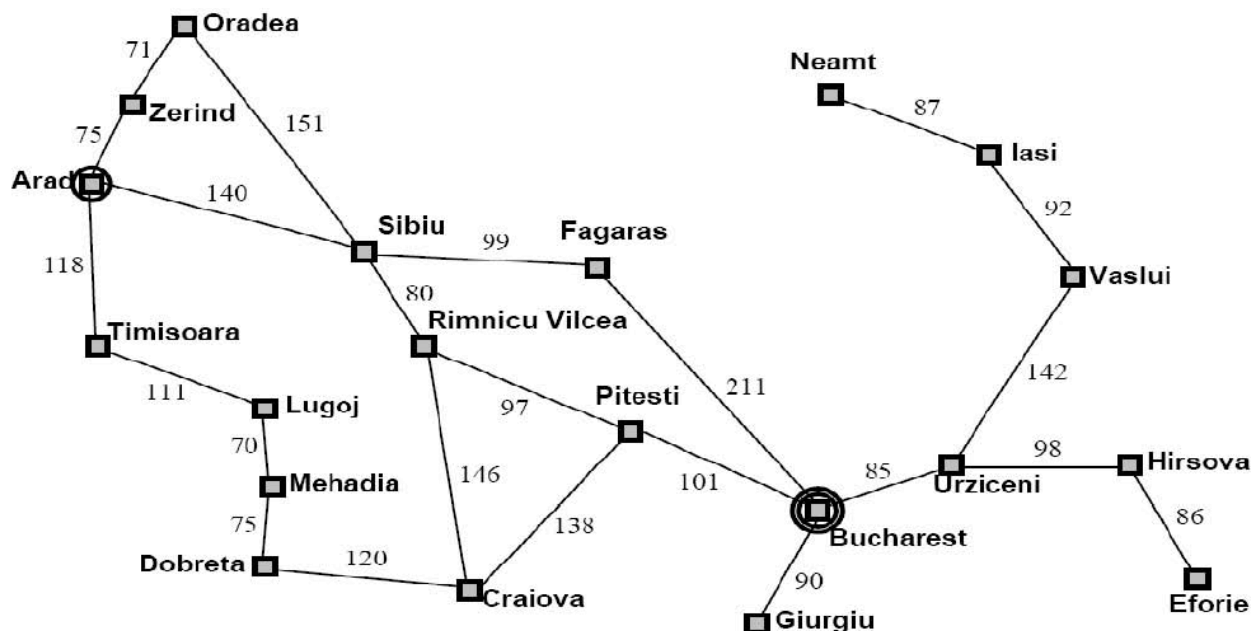
• 盲目搜索

- 宽度优先搜索
- 一致代价搜索
- 深度优先搜索
- 深度受限搜索
- 迭代加深搜索
- 双向搜索

• 启发式搜索

- A*搜索算法
- IDA*搜索算法

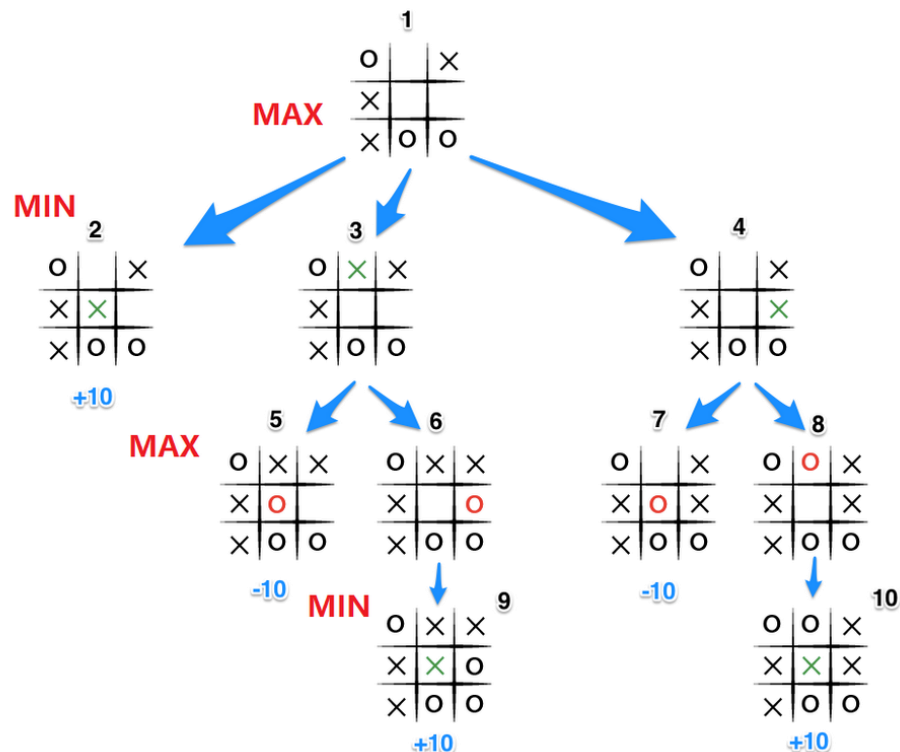
Currently in Arad, need to get to Bucharest





Lab10 博弈树搜索

- 零和博弈
- MiniMax策略
- Alpha-beta剪枝



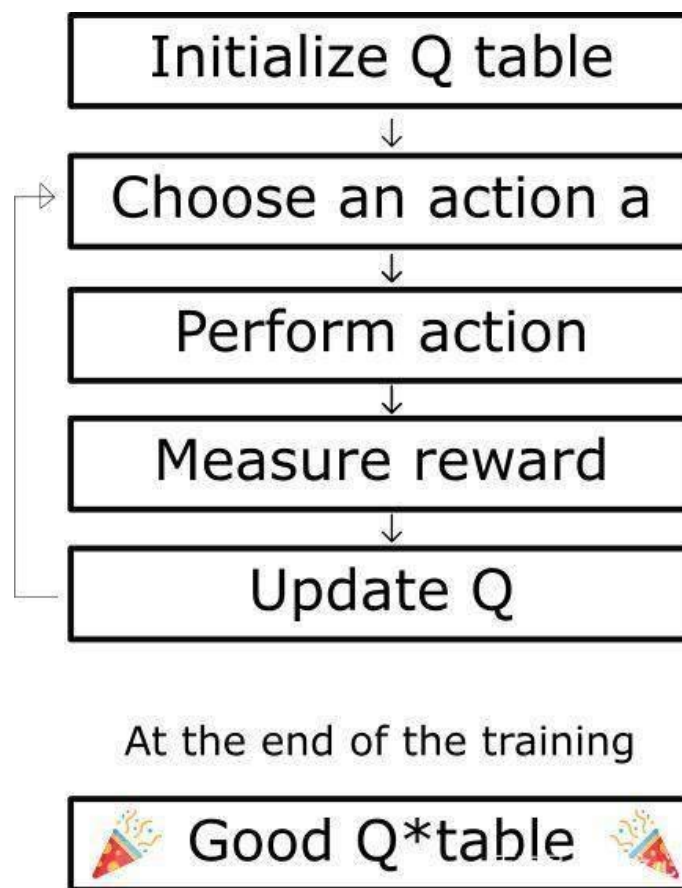


Lab11 Q-Learning

- 一种蒙特卡洛方法，通过建立Q-table来进行策略判断。

$$Q(s, a) = R(s, a) + \gamma \max_{a'} \{Q(s', a')\}$$

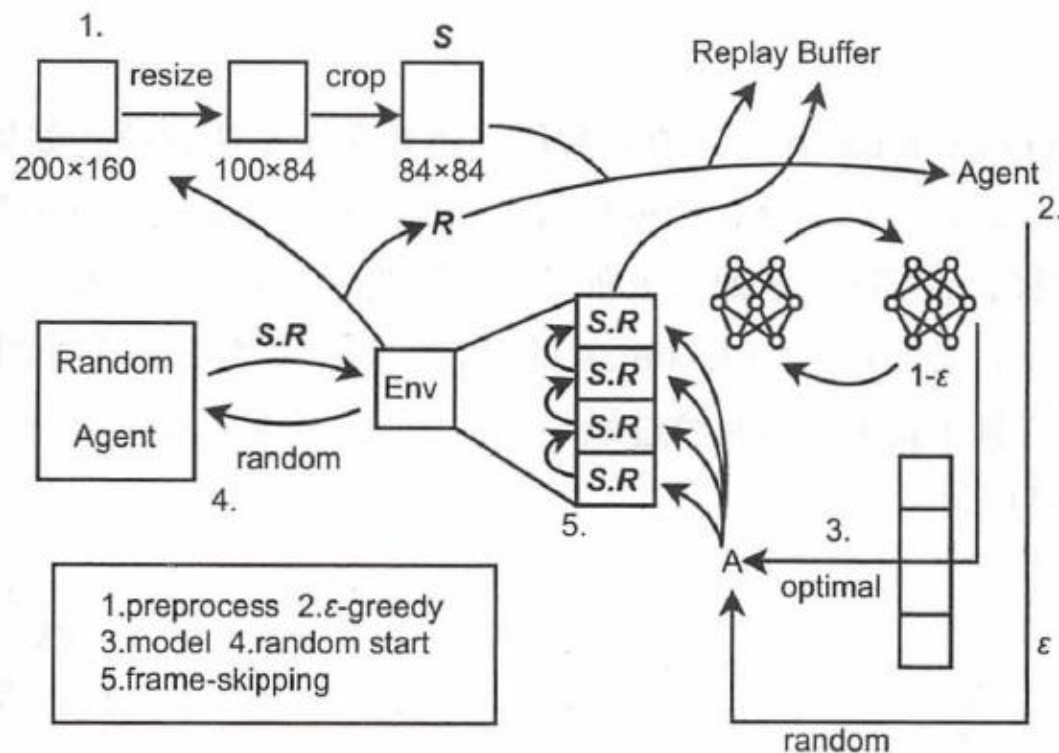
- 只能在有限状态空间上进行。





Lab12 DQN与Policy Gradient

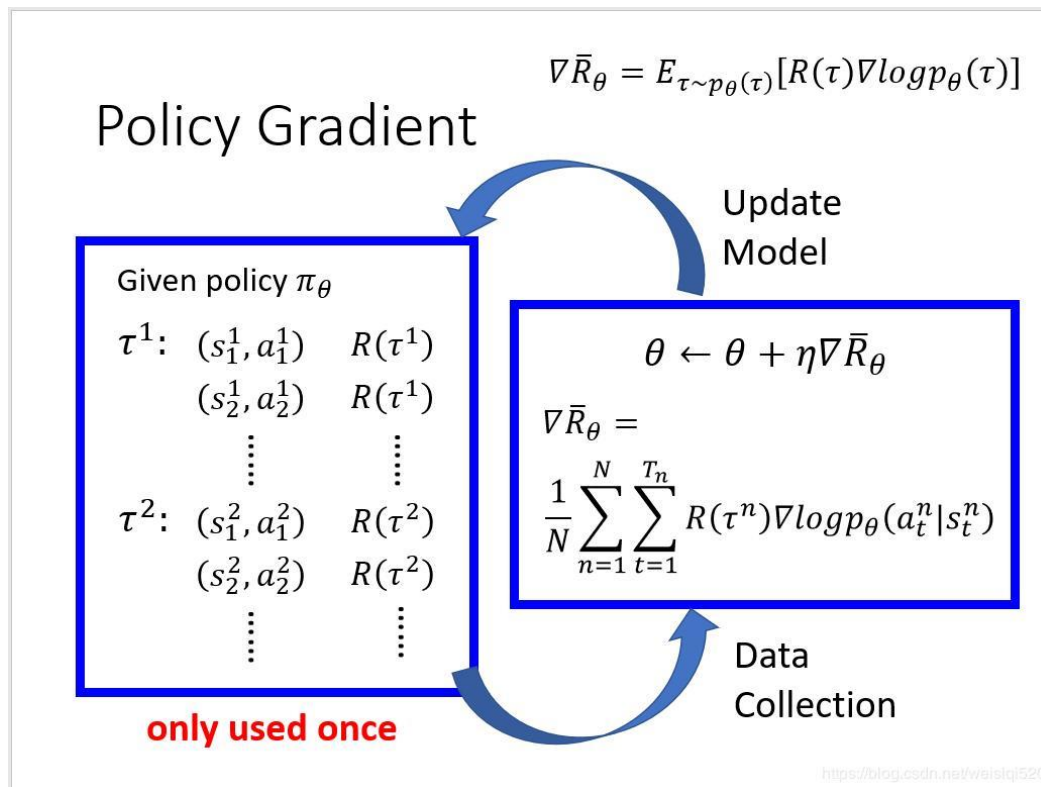
- 值函数近似：使用一个机器学习模型来拟合值函数Q
- DQN 结合了Q-Learning的价值估计方法和深层模型较强的拟合效果。





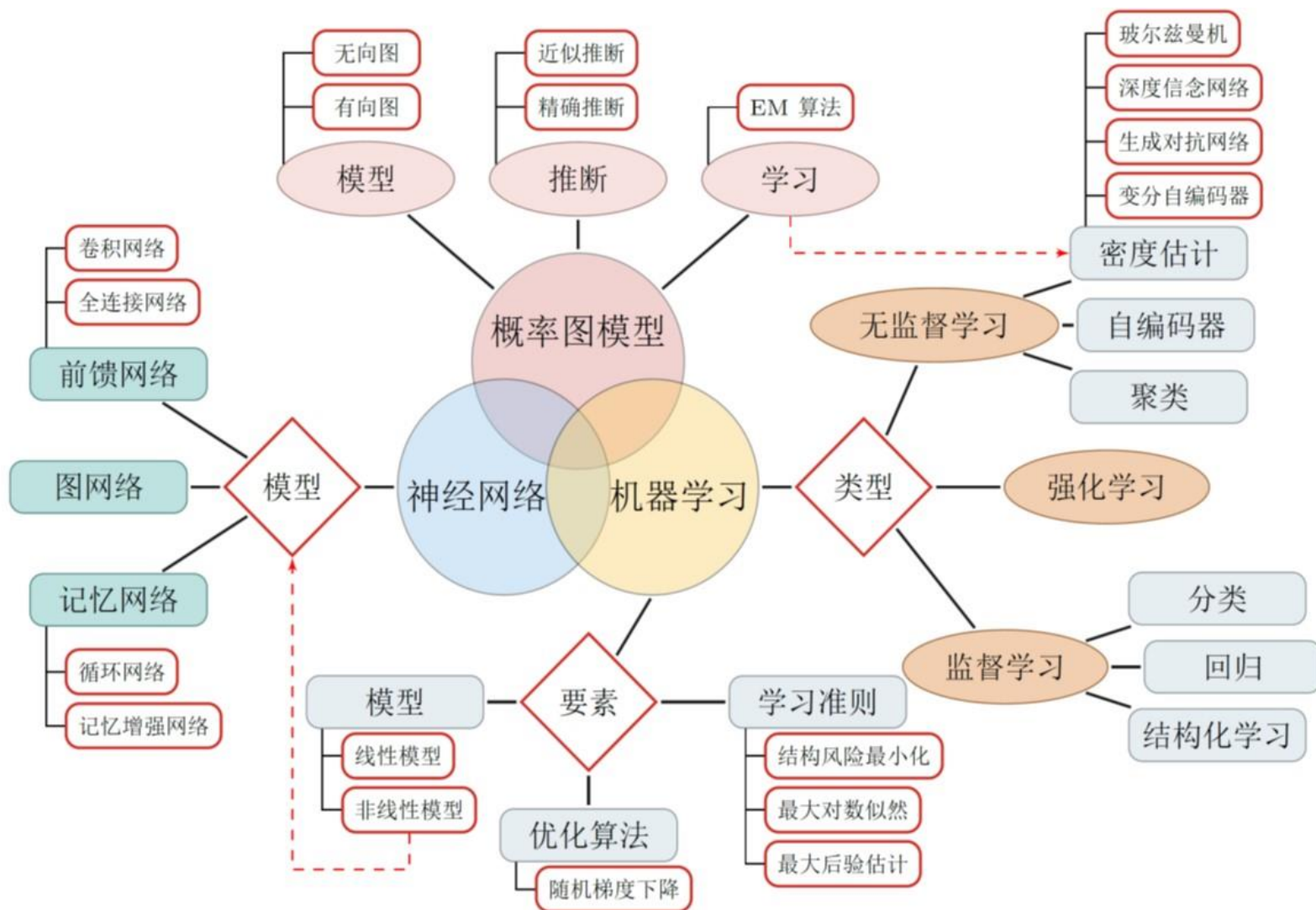
Lab12 DQN与Policy Gradient

- Value-based的方法缺点：
 - 对连续动作的处理能力不足
 - 无法解决随机策略问题
- Policy-based：在策略空间直接搜索来得到最佳策略。
- 参数化的策略能够处理连续状态和动作，可以直接学出随机性策略。





学习模式



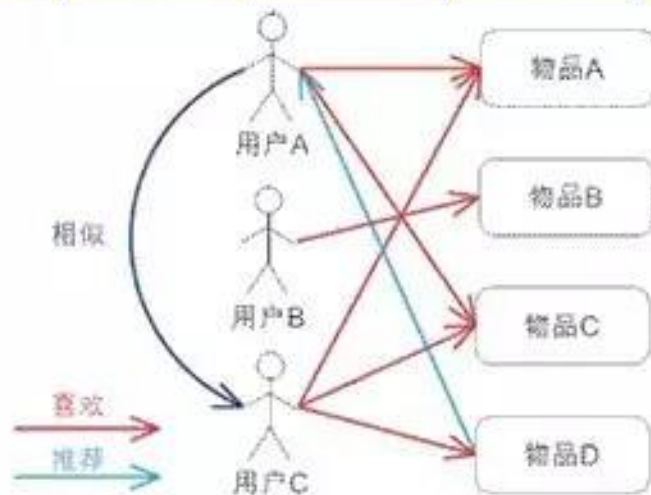


网易云音乐



- 基本算法基于协同过滤
(Collaborative filtering)
- 使用神经网络算法解决CF中的冷启动问题。
- 初期辅以一定的人工

用户/物品	物品A	物品B	物品C	物品D
用户A	√		√	推荐
用户B		√		
用户C	√		√	√





不存在的人

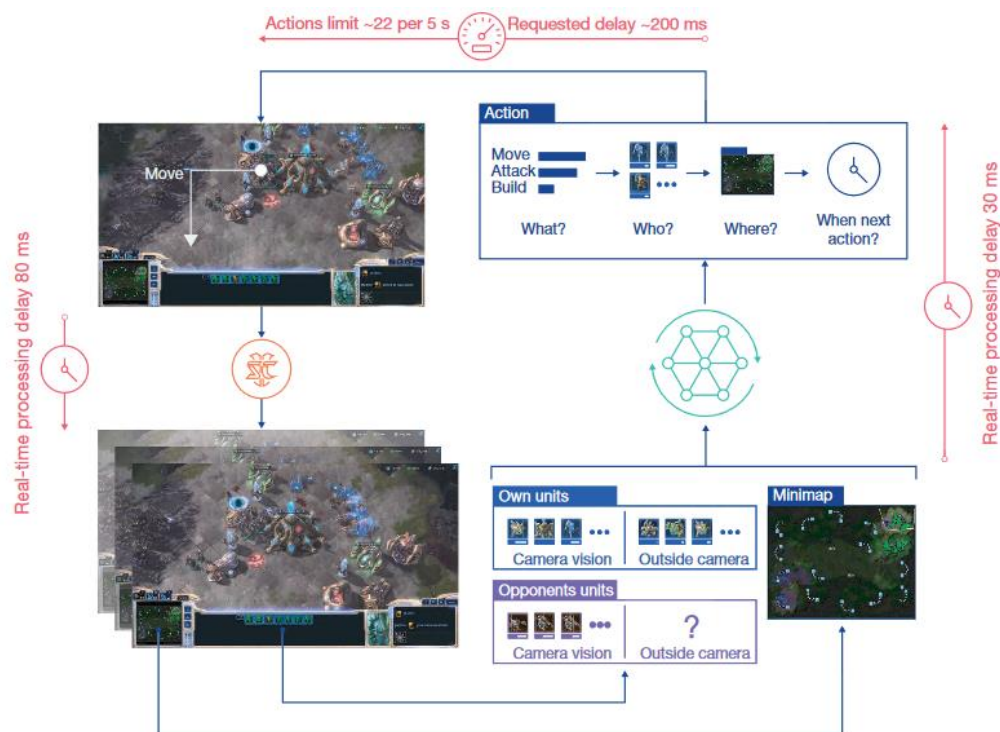
- 使用对抗生成网络生成人脸数据集
- <https://thispersondoesnotexist.com/>





AlphaStar

- 状态空间约为 10^{26} 的量级
- 使用LSTM来处理观察的时序序列
- 强化学习部分基于 advantage actor-critic





期末小组对弈流程

第一阶段：积分赛

① 将队伍分为四组，首先组内进行两两对弈，每胜利一局加上一分，失败一局不加分。小组内积分最高的队伍进行决赛对弈。

② 具体流程：按照分组情况，分为四个小组，每个TA负责一个小组，分好组后，小组按照规定的座位号进行就坐，当胜负分出后，**保持界面，举手示意TA，TA负责登记后，此次比赛才算结束。**

第二阶段淘汰赛（决赛）

① 四支进入决赛的队伍，随机抽签进行两场对弈比赛，胜利的两支队伍角逐第一二名，失败的两支队伍角逐第三四名。

② 具体流程：两场比赛同时进行，在第一排各两个TA的监督下进行，胜负分出后，TA进行记录。



期末小组对弈流程

每场比赛细则：

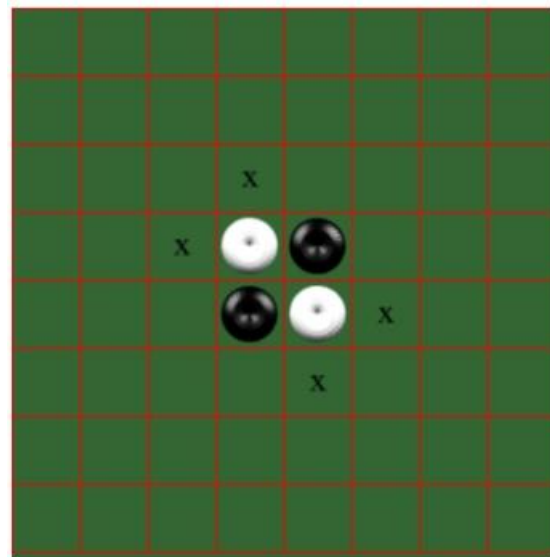
每组按照比赛流程，单盘棋局比赛时间一共为13分钟，然后开始更换对弈目标，休息调整，每15分钟开始一局。

获胜条件：

①当场上的棋子全为一个颜色时，该颜色方胜利。②棋局无空白位置时，颜色棋子多者获胜。③十三分钟到时，若后手下完，则按照后手下完的局面统计棋子数量；若先手刚下完，则允许后手再下一步，统计之后的局面。

实现要求：

- ① 棋盘初始化为下图所示，一盘棋局黑棋先手。
- ② 每次下一步棋后（无论是对手还是自己下），更新输出棋盘黑白棋子两方数量。
- ③ 实现两种模式，电脑先手和人类先手，与五子棋相同的对弈流程。



初始棋局



期末小组对弈流程

评分细则 (rank占期末项目的20分)

- ① 前四名队伍，第一名20分，第二三名19分，第四名18分。
- ② 未进入决赛队伍，基础分为12分，按照小组赛积分，每积两分在基础分上加一分。即如果小组积分赛积分为1或2分，则rank分数为13；若积分为7或8分，则rank分数为16。