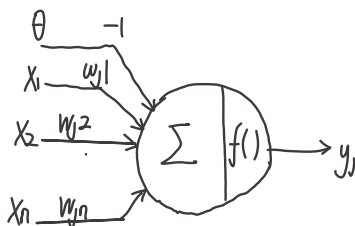
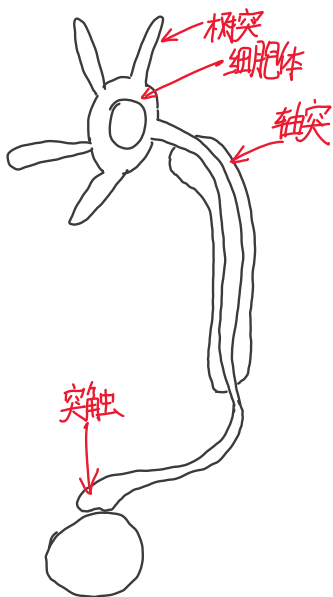


生物神经元和人工神经元对比



θ 神经元单位的偏置

w_i 神经元间连接强度 (激发状态取正数, 抑制状态取负数)

n 输入神经元数

y 神经元输出

t 时间

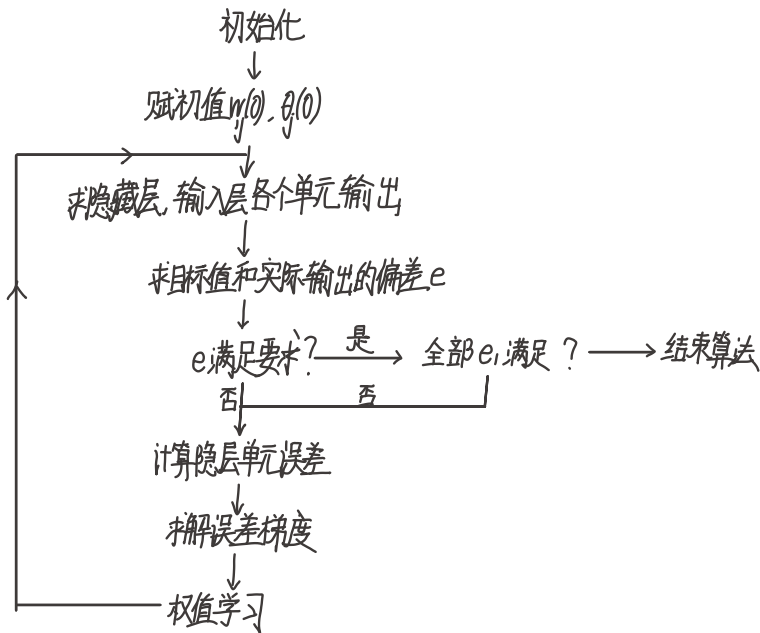
$f()$ 输出变换函数 / 激励函数

$$\begin{cases} u_j(t) = \sum_{i=1}^n w_{ji} x_i - \theta_j \\ y_j(t) = f\left(\sum_{i=1}^n w_{ji} x_i - \theta_j\right) \end{cases}$$

样例学习 概率模型学习 深度学习 强化学习

BP学习算法流程图/步骤

- 1 初始化, 对所有连接权重和偏置进行赋值随机小值
- 2 从 N 组样本中选取一组样本 X , 将信息输入 BP 网络中
- 3 正向传播, 计算各层节点的输出 y'
- 4 计算网络实际输出和期望输出的误差 $e_i = y - y'$
- 5 反向传播, 向减小误差方向调整网络连接权值
- 6 取出另一组样本重复 (2) ~ (5), 直到 N 组输入输出样本误差达到要求



BP网络优缺点

1 有良好的逼近特性 2 有较强的泛化能力 3. 有较好的容错性

1 收敛速度慢 2 可能收敛到局部极值 3. 难以确定隐层和隐层结点数

卷积神经网络和人工神经网络的区别

后者 1 输入信息是向量

2 神经元一维排列

3 由输入层、隐藏层、输出层组成

4 同层神经元之间不连接，每个神经元和前一层所有神经元全连接

前者 1 输入信息是图像

2 神经元三维排列

3 由输入层、卷积层、池化层、激活层、全连接层、输出层组成

4 卷积层的每个神经和前一层以感受野形式连接，仅部分连接/稀疏连接

卷积层的特性 局部连接、参数共享

优点：减小参数量、节约计算成本

池化层的作用：减小数据量，减轻过拟合

全连接层的作用 作为分类器

激活层的作用 非线性拟合，使网络拟合出的特征更复杂

CNN的优点：

- 1 隐式的从训练数据中学习，避免了显式的特征抽取
- 2 权值共享降低了网络学习复杂性，使网络可以并行学习
- 3 在语音识别和图像处理方面有优势
- 4 采用时间和空间的下采样结构，可以获得某种程度的位移、尺度、形变的鲁棒性

缺点

- 1 结构参数随意性、试凑性强，不能保证拓扑参数收敛
- 2 没有反馈结构
- 3 训练样本要求海量

智能学习概论

智能体的问题生成器、评判标准、学习元素可以作用机器学习

监督学习 < 回归
 < 分类

偏差大 \rightarrow 欠拟合 方差大 \rightarrow 过拟合

偏差 - 方差权衡

超参数

有多模型可供选择时，模型选择可以在验证集上交叉验证来选出较好的

训练目标是最小化验证集上的损失函数

计算学习理论

决策树 可以表示任何布尔函数，基于信息增益的启发式算法可以找到他

线性回归

感知机：带有硬阈值的线性分类器，不能处理线性不可分的数据

逻辑斯蒂回归：将感知机硬阈值替换为 函数定义的软阈值，
可以处理带噪声的线性不可分数据集

非参数模型：最近邻和局部加权回归

支持向量机：寻找带有最大边距的线性分离器，改进分类的泛化表现

核方法 对于线性不可分的原数据，可以隐式地把输入数据投影到
可能存在线性分离器的高维空间中

集成方法 自助聚合法、自适应提升法、随机森林、堆叠法

在线学习：对于不断变化的数据分布，可以通过聚集专家意见达到任意程度接近
最好专家表现

线性形式下有参数模型
对偶形式下无参数模型

概率学习模型

14 完全数据学习

数据 + 假设

- 1 贝叶斯学习方法 学习表示为概率推理 | 利用观测值对先验分布做更新, 呈现了奥卡姆剃刀原理, 但难以处理更复杂的假设空间
- 2 MAP 学习 (最大后验) 选择给定数据下可能性最大的假设。同样使用假设先验分布, 但比上一方法更易处理
- 3 MLE (最大似然) 学习 选择使得数据似然最大的假设, 等价于使用均匀分布的最大后验学习
→ 假定特征相互独立
优点: 可以很好处理含噪声或缺失数据情况: 计算简单
缺点: 独立性假设实际中不成立, 假设会导致对某些概率过于自信的估计 (接近 0/1)
- 4 朴素贝叶斯学习 (隐变量学习)
- 5 期望最大化算法: 处理部分变量被隐藏的情况, 可以找到局部最大似然解
(实例: 高斯混合模型无监督聚类、贝叶斯网络学习、隐马尔可夫模型的学习)
- 6 结构空间的离散搜索 → 模型选择 贝叶斯网络学习
- 7 非参数模型 最近邻方法、核方法, 特点是用一些数据点集合表示某一分布, 因此其参数数量随着训练集的增大而增大

隐变量可以减少确定贝叶斯网络所需的参数个数和所需学习的参数个数

概率模型到生成模型: 能解释观测数据, 找到最可能生成数据的模型

传统概率模型 找到概率分布 $p_{\theta}(x) \quad \theta^* = \underset{\theta}{\operatorname{argmax}} p_{\theta}(x)$ [概率建模]

生成式概率模型 概率建模 + 采样并生成数据

深度学习

LSTM

- 1 遗忘门 f_t 决定了记忆单元中每个元素是被记生还是遗忘
- 2 输入门 i_t 决定了记忆单元中每个元素是否与来自当前时间步的输入向量新信息做加法
- 3 输出门 o_t 决定了记忆单元中每个元素是否被转移到短期记忆 z_t

无监督学习 PCCA (概率主成分分析)、自编码器、变分自编码器、深度自回归模型
GAN (生成对抗网络)

迁移学习

强化学习

智能体 3 要素: 感知、决策和奖励
关键

被动强化学习分类

(回报)

1 基于模型的强化学习 智能体配备环境的转移模型 $P(s'|sa)$, 并学习效用函数 $U(s)$

2 无模型强化学习 智能体学习动作效用函数 $Q(s,a)$ 或学习一个策略 $\pi(s)$

方法论分类

没有效用值, 满足贝尔曼方程的约束

直接效用估计
时序差分方法

自适应动态规划

学习效用函数方法

$$U(s) = \sum_{s'} P(s'|s, \pi(s)) [R(s, \pi(s), s') + \gamma U(s')]$$

主动和被动强化学习区别 后者有一个固定策略决定其行为, 而主动强化学习中智能体要自主决定采取的动作

强化学习改进思路 奖励设计和分层强化学习, 用于处理奖励稀少且需要长动作序列才能生效的情况

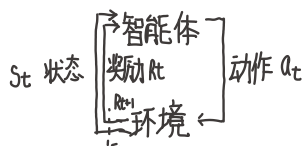
策略搜索 直接对策略的表示进行操作, 跟踪观测到的表现进行改进
观测专家行为进行学徒学习 \rightarrow 无法获得正确奖励函数时的解决方案

模仿学习 问题转换为从专家的状态-动作中进行学习的监督学习

逆强化学习 从专家行为中推断有关奖励函数的信息

强化学习概念 智能体和世界进行互动并定期收到反映其表现的奖励的学习方法

强化学习基本结构 智能体、环境、状态、动作、奖励、策略、目标



状态转移 $P_t(s'|s,a)$ [随机, 确定]

策略 $\pi(a|s)$ [随机, 确定, 贪婪]

奖励 $r_t = r(s_t, a_t)$ [确定, 不确定]

回报函数 $U_t = r_t + \gamma^0 r_{t+1} + \gamma^1 r_{t+2} \dots$ (从 t 时刻到终止状态的奖励总和)

动作价值 $Q_t(s_t, a_t)$ (状态 s 执行动作 a 后对后续回报 U 期望)
状态价值 $V_t(s_t)$

主动强化学习

贪心智能体的问题 只关注动作提供的奖励，忽略了结果状态中以感知形式提供的信息，智能体必须在利用当前最佳动作最大化其短期奖励和探索过去未知状态以获得可能导致策略改变的信息间做权衡 (老虎机问题)

GLIE方案 每个状态下对每个尝试选取任意多次，避免错过最优动作

$$U^{\pi}(s) \leftarrow \max_a f(\sum_p(s'|s,a)[R(s,a,s') + \gamma U^{\pi}(s')], N(s,a))$$

f 探索函数，作用是驱动智能体去访问没有探索的部分

安全搜索

Q学习更新Q值使用状态 s' 最佳动作 a 对应的Q值进行更新

时序差分Q学习

SARSA

使用实际采取的动作 a' 对Q值进行更新

Q学习 离策略学习算法，Q值回答

停止当前

选择最佳动作策略

SARSA 同策略学习算法，Q值回答

坚持

动作在当前状态下的价值

(DRL)

深度强化学习 使用深度网络作为效用函数或Q函数近似器的强化学习方法

↳ 分类 离散的 连续的

传统强化学习缺点 1 状态空间随问题规模指数增长 2 探索效率低

3 难以捕捉复杂模式

深度强化学习缺点 1. 训练成本高 2 训练不稳定 3 解的质量不稳定

↓

梯度消失/爆炸

泛化能力好，端到端优化

优化策略和产生行为策略相同

↓

$$Q(s,a) \leftarrow Q(s,a) + \alpha [R + \gamma Q(s',a') - Q(s,a)]$$

SARSA 更新Q值公式

$$Q(s,a) \leftarrow Q(s,a) + \alpha [R + \gamma \max_{a'} Q(s',a') - Q(s,a)]$$

Q学习更新Q值公式

↑ 优化策略和产生行为策略不同

奥卡姆剃刀原理：如无必要，勿增实体；当一个现象有多种可能的假说时，最简单直观的假说往往是正确的，学习系统也倾向于学习简单假设

$$D \text{ 为数据, } M \text{ 为假说, 故 } P(M|D) = \frac{P(D|M) P(M)}{P(D)}$$

$$\text{若各假说先验概率相等, 则 } M^* = \underset{M}{\operatorname{argmax}} P(M|D) = \underset{M}{\operatorname{argmax}} P(D|M)$$

复杂假说会把概率分散到更多可能的数据上，从而对特定观测数据预测概率降低。贝叶斯推断基于上述公式惩罚过度复杂模型，从而选择最有效简单模型

贪心智能体为什么策略次优

- 1 学习到的模型和真实环境不一样
- 2 忽略了动作不仅提供奖励还以感知方式提供信息，智能要在选择最动作以最大化短期奖励和探索过去未知状态以获得可能导致策略改变的信息间做平衡了

