

Part III-B: Artificial Intelligence Outline

Lecture by 熊庆宇

Note by THF

2024 年 11 月 1 日

目录

0.1 遗传算法	1
0.2 蚁群算法	2
0.3 机器学习	3
0.3.1 监督学习	4

Lecture 11

10.28

0.1 遗传算法

Notation. 如何模拟物种繁殖:

1. 种群中选择两个个体
2. 随机确定编码序列断裂点
3. 交换编码片段

基因发生突变的概率称为遗传算法的**突变算子**

Notation. 如何模拟竞争与选择: **适应度**

Example. 求函数 $f(x) = x^2$ 的最大值, $x \in [0, 31], x \in \mathbb{Z}$

使用 5 个二进制码表示取值: $0 \sim 31 \Rightarrow 0b00000 \sim 0b11111$

定义 32 条染色体:

表 1: 染色体

X_o	X_b
0	0b00000
1	0b00001
...	...
31	0b11111

假设初始种群数量 $N = 4$ ，随机产生 20 位的二进制串，每 5 个一组，得到 4 个初始个体
如：0b**001101**0010**100110**1010

适应度函数为给定问题 $f(x) = x^2$

表 2: 选择算子

编号	个体编码	个体	适应度	$\frac{f}{\sum f}$	$\frac{4f}{\sum f}$	生存数
S_1	00110	6	36	0.043	0.175	0
S_2	10010	18	324	0.394	1.578	2
S_3	10011	19	361	0.439	1.759	2
S_4	01010	10	100	0.121	0.487	0
适应度总和			821	平均适应度		205.25

得到第一代种群：(0b10010, 0b10011), (0b10010, 0b10011)：避免近亲相交

令交叉概率 $P_s = 1$ ，变异概率 $P_m = 0.01$ （每五代变异一个基因），通过生存数生成新的种群，配对后随机选择断裂点位交叉配对，完成第一代遗传：第二代种群的适应度更高

表 3: 第一次交叉

交叉前	交叉后	个体	适应度	生存数
0b 10010	0b10011	19	361	1
0b 10011	0b10010	18	324	1
0b 10010	0b10011	19	361	1
0b 10011	0b10010	18	324	1
适应度总和			821 \Rightarrow 1370	

生成第五代种群后，通过变异算子随机挑选一个基因进行改变

经过数代遗传后，种群趋于稳定，适应度不再提升： $X = 31$

0.2 蚁群算法

Notation. 蚂蚁的智能程度非常低，单个觅食随机性很大；但组合成群体后可以完成复杂的任务，且可以适应环境变化

Notation. 蚂蚁依靠**信息素**寻找最短路径

信息素：蚂蚁自身释放的易挥发的物质

在该道路上经过的蚂蚁越多，信息素浓度越高；浓度越高，这条道路就越容易被选择

信息素会随着时间推移而消散

Notation. 正反馈机制：

在寻找到较短路径后，蚂蚁释放的信息素会增加蚂蚁选择该路径的概率，同时后续的蚂蚁释放的信息素会进一步加强信息素浓度

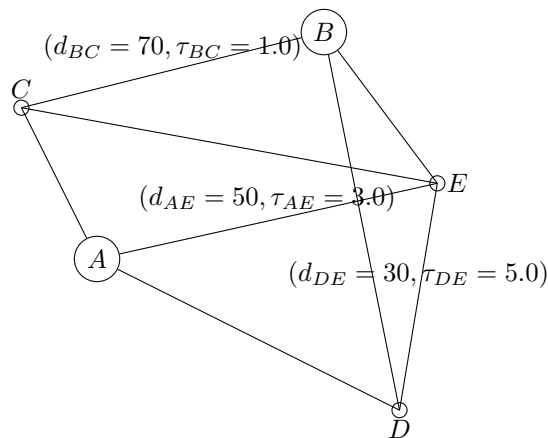
模拟蚂蚁觅食的 4 个抽象部分:

模拟蚂蚁

- 相同目标, 相同速度运动
- 在到达目的地之前不回头、不转圈
- 根据相同的原则释放信息素、选择路径
- 记得自己走过的路径长度
- 种群中的个体数量不变

模拟地图

具有 N 个节点的全连通图, 任意两点 X, Y 之间的距离 d_{XY} 设为已知, 具有明确的起点与终点



Example. 旅行商问题

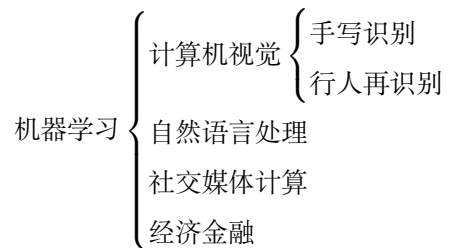
0.3 机器学习

Definition. 学习:

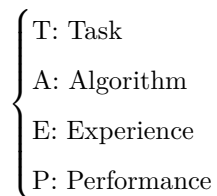
- 系统改进其性能的过程 (西蒙)
- 获取知识的过程 (专家系统)
- 技能的获取 (心理学家)
- 事物规律的发现过程

Lecture 12

10.31



机器学习的四个部分：



Example. 人脸识别：

A：线性回归

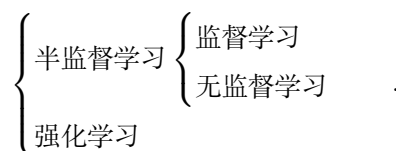
E：以标定身份的人脸图片数据

P：人脸识别准确率

机器学习的基本过程

从给定的数据中学习规律 → 学习方法，建立模型 → 预测 → 测试匹配度

机器学习分类



0.3.1 监督学习

Definition. 根据已知的输入和输出训练模型，预测未来输出

监督学习的数据存在样本标签，有训练集和测试集

Example. 学习书籍内容，设定标签：艺术/政治/科学等，找出训练文字和标签的映射关系

Notation. 分类方法： K -nearest neighbour，决策树，支持向量机，朴素贝叶斯

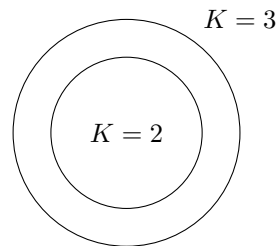
回归方法：线性、树、支持向量回归，集成方法

K-nearest neighbor

基本思路

找到与新输入的待预测样本最临近的 K 个样本，判断这 K 个样本中绝大多数的所属类别作为分类结果输出

条件：已经具有较大的样本量



Notation. KNN 算法的基本要素：距离度量、 K 值、分类决策规则

距离度量

Notation. KNN 算法能够分类：特征空间内的样本点之间的距离能够反映样本特征的相似程度

设有两个样本点 $\mathbf{x}_i, \mathbf{x}_j$ ，以 n 维向量空间作为特征空间，将这两个点表示为：

$$\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}.$$

$$\mathbf{x}_i = (x_i^1, x_i^2, \dots, x_i^n)^T.$$

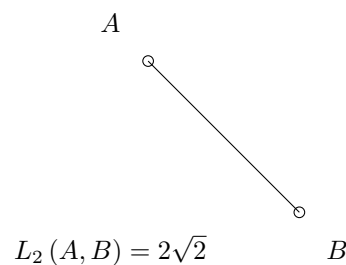
$$\mathbf{x}_j = (x_j^1, x_j^2, \dots, x_j^n)^T.$$

特征点之间的距离定义为：

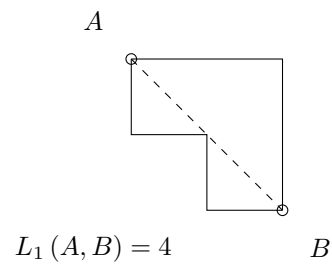
$$L_p(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{l=1}^n |x_i^l - x_j^l|^p \right)^{\frac{1}{p}}.$$

Example. 代入 $p = 2$ ，易得 $L_2(\mathbf{x}_i, \mathbf{x}_j)$ 为平面上两点间的距离公式，该距离又称为欧氏距离：

$$L_2(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2}.$$



代入 $p = 1$: $L_1(x_i, x_j)$ 称为曼哈顿距离:



K 值的选择

使用交叉验证方法确定最合适的 K 值