一、知识表示

1.产生式表示法：

①确定性规则知识的产生式表示：IF P THEN Q 或者 P→Q

例如：IF 动物会飞 AND 会下蛋 THEN 该动物是鸟

②不确定性规则知识的产生式表示：IF P THEN Q（置信度）或者 P→Q（置信度）

例如：IF 发烧 THEN 感冒 (0.6)

③确定性事实性知识的产生式表示：三元组表示（对象，属性，值）或者（关系，对象 1，对象 2）

例如：老李年龄是 40 岁：(Li,age,40)；老李和老王是朋友：(friend,Li,Wang)

④不确定性事实性知识的产生式表示：四元组表示（对象，属性，值，置信度）或者（关系，对象 1，对象 2，置信度）

例如：老李年龄很可能是 40 岁：(Li,age,40,0.8)；老李和老王不大可能是朋友：(friend,Li,Wang,0.1)

2.框架表示法

<框架名>

槽名 1：侧面名$_{11}$ 侧面值$_{111}$,…,侧面值$_{11P1}$

…

槽名 n：侧面名$_{1m}$ 侧面值$_{1m1}$,…,侧面值$_{1mPm}$

约束：约束条件 1

…

约束条件 n

特点：结构性、继承性、自然性

【例题】某年某月某日，某地发生 6.0 级地震，若以膨胀注水孕震模式为标准，则三项地震前兆中的波速比为 0.45，水氡含量为 0.43，地形改变为 0.60。

框架名：<地震>

地点：某地

日期：某年某月某日

震级：6.0

波速比：0.45

水氡含量：0.43

地形改变：0.60

<地震>
地点 某地
日期

3.知识图谱：是一种互联网环境下的知识表示方法，目的是提高搜索引擎的能力，改善用户的搜索质量以及搜索体验，三元组是知识图谱的通用表示方法

• 定义：知识图谱，又称科学知识图谱，用各种不同的图形等可视化技术描述知识资源及其载体，挖掘、分析、构建、绘制和显示知识及它们之间的相互联系。

逻辑结构：模式层、数据层

（实体 1-关系-实体 2）：中国-首都-北京

（实体-属性-属性值）：北京-人口-2069 万

二、模糊集合

1.定义

论域：所讨论的全体对象，用 U 等表示。

元素：论域中的每个对象，常用 a,b,c,x,y,z 表示。

集合：论域中具有某种相同属性的确定的、可以彼此区别的元素的全体，常用 A,B 等表示。

元素 a 和集合 A 的关系：a 属于 A 或 a 不属于 A，即只有两个真值“真”和“假”。

模糊逻辑给集合中每一个元素赋予一个介于 0 和 1 之间的实数，描述其属于一个集合的强度，该实数称为元素属于一个集合的隶属度，集合中所有元素的隶属度全体构成集合的隶属函数。

$$A = \{(x, \mu_A(x)), x \in X\}$$

$A = \{(x, \mu_A(x)), x \in X\}$

其中，$\mu_A(x)$是元素 x 属于模糊集 A 的隶属度，X 是元素 x 的论域

2.表示方法

①Zadeh 表示法：

论域是离散且元素数目有限：$A = \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + \cdots + \frac{\mu_A(x_n)}{x_n} = \sum_{i=1}^{n} \frac{\mu_A(x_i)}{x_i}$

或$A = \{\frac{\mu_A(x_1)}{x_1}, \frac{\mu_A(x_2)}{x_2}, \ldots, \frac{\mu_A(x_n)}{x_n}\}$

②序偶表示法：

$$A = \{(\mu_A(x_1), x_1), (\mu_A(x_2), x_2), \ldots, (\mu_A(x_n), x_n)\}$$

③向量表示法：

$$A = \{\mu_A(x_1), \mu_A(x_2), \ldots, \mu_A(x_n)\}$$

3.求 A 到 B 的模糊关系 R

$$A = \frac{1.0}{a_1} + \frac{0.8}{a_2} + \frac{0.5}{a_3} + \frac{0.2}{a_4} + \frac{0.0}{a_5}$$

$$B = \frac{0.7}{b_1} + \frac{1.0}{b_2} + \frac{0.6}{b_3} + \frac{0.0}{b_4}$$

求解：$R = A'B = \mu_A^T \circ \mu_B = \begin{bmatrix} 1.0 \\ 0.8 \\ 0.5 \\ 0.2 \\ 0.0 \end{bmatrix} \circ [0.7 \quad 1.0 \quad 0.6 \quad 0.0]$

$R = A'B$

$\wedge$ 取较小的 $= \mu_A^T \circ \mu_B$

$$R = \begin{bmatrix} 1.0\wedge0.7 & 1.0\wedge1.0 & 1.0\wedge0.6 & 1.0\wedge0.0 \\ 0.8\wedge0.7 & 0.8\wedge1.0 & 0.8\wedge0.6 & 0.8\wedge0.0 \\ 0.5\wedge0.7 & 0.5\wedge1.0 & 0.5\wedge0.6 & 0.5\wedge0.0 \\ 0.2\wedge0.7 & 0.2\wedge1.0 & 0.2\wedge0.6 & 0.2\wedge0.0 \\ 0.0\wedge0.7 & 0.0\wedge1.0 & 0.0\wedge0.6 & 0.0\wedge0.0 \end{bmatrix} = \begin{bmatrix} 0.7 & 1.0 & 0.6 & 0.0 \\ 0.7 & 0.8 & 0.6 & 0.0 \\ 0.5 & 0.5 & 0.5 & 0.0 \\ 0.2 & 0.2 & 0.2 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

▪**例3.3 设模糊集合**$X = \{x_1, x_2, x_3, x_4\}, Y = \{y_1, y_2, y_3\}, Z = \{z_1, z_2\}$

$Q \in X \times Y, R \in Y \times Z, S \in X \times Z$，求 S。

$$Q = \begin{bmatrix} 0.5 & 0.6 & 0.3 \\ 0.7 & 0.4 & 1 \\ 0 & 0.8 & 0 \\ 1 & 0.2 & 0.9 \end{bmatrix} \qquad R = \begin{bmatrix} 0.2 & 1 \\ 0.8 & 0.4 \\ 0.5 & 0.3 \end{bmatrix}$$

解：

$$S = Q \circ R = \begin{bmatrix} 0.5 & 0.6 & 0.3 \\ 0.7 & 0.4 & 1 \\ 0 & 0.8 & 0 \\ 1 & 0.2 & 0.9 \end{bmatrix} \circ \begin{bmatrix} 0.2 & 1 \\ 0.8 & 0.4 \\ 0.5 & 0.3 \end{bmatrix}$$

$$= \begin{bmatrix} (0.5\wedge0.2)\vee(0.6\wedge0.8)\vee(0.3\wedge0.5) & (0.5\wedge1)\vee(0.6\wedge0.4)\vee(0.3\wedge0.3) \\ (0.7\wedge0.2)\vee(0.4\wedge0.8)\vee(1\wedge0.5) & (0.7\wedge1)\vee(0.4\wedge0.4)\vee(1\wedge0.3) \\ (0\wedge0.2)\vee(0.8\wedge0.8)\vee(0\wedge0.5) & (0\wedge1)\vee(0.8\wedge0.4)\vee(0\wedge0.3) \\ (1\wedge0.2)\vee(0.2\wedge0.8)\vee(0.9\wedge0.5) & (1\wedge1)\vee(0.2\wedge0.4)\vee(0.9\wedge0.3) \end{bmatrix}$$

$$= \begin{bmatrix} 0.6 & 0.5 \\ 0.5 & 0.7 \\ 0.8 & 0.4 \\ 0.5 & 1 \end{bmatrix}$$

其中，$\wedge$运算：两个取较小者；$\vee$ 运算：两个取较大者

矩阵 左出行，右出列

最后每行出最大

4.根据模糊关系 R 求推理结果

[04, 10], [0, 0 1, 0 0]

▪ 当输入： $A' = 0.4/a_1 + 0.7/a_2 + 1.0/a_3 + 0.6/a_4 + 0.0/a_5$

$$B' = A' \circ R = \begin{bmatrix} 0.4 \\ 0.7 \\ 1.0 \\ 0.6 \\ 0.0 \end{bmatrix}^T \circ \begin{bmatrix} 0.7 & 1.0 & 0.6 & 0.0 \\ 0.7 & 0.8 & 0.6 & 0.0 \\ 0.5 & 0.5 & 0.5 & 0.0 \\ 0.2 & 0.2 & 0.2 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix} = (0.7, 0.7, 0.6, 0.0)$$

则： $B' = 0.7/b_1 + 0.7/b_2 + 0.6/b_3 + 0.0/b_4$

0.4 0.7 0.5 0.2 0
0.4 0.7 0.5 0.2 0
0.4 0.8 0.5

5.模糊决策

①最大隶属度法：找出隶属度（分子）最大的一项作为结果（若有多项，则求平均）

②加权平均判决法：Σ隶属度（分子）×论域（分母）/Σ隶属度（分子）

【例题】设有模糊控制规则：

“如果温度低，则将风门开大”。设温度和风门开度的论域为{1，2，3，4，5}。

“温度低”和“风门大”的模糊量：

“温度低”=1/1+0.6/2+0.3/3+0.0/4+0/5；“风门大”=0/1+0.0/2+0.3/3+0.6/4+1/5

已知事实“温度较低”，可以表示为：“温度较低”=0.8/1+1/2+0.6/3+0.3/4+0/5

试用模糊推理确定风门开度。

（1）确定模糊关系 *R*

$$R = \begin{bmatrix} 1.0 \\ 0.6 \\ 0.3 \\ 0.0 \\ 0.0 \end{bmatrix} \circ [0.0 \quad 0.0 \quad 0.3 \quad 0.6 \quad 1.0]$$

$$= \begin{bmatrix} 0.0 & 0.0 & 0.3 & 0.6 & 1.0 \\ 0.0 & 0.0 & 0.3 & 0.6 & 0.6 \\ 0.0 & 0.0 & 0.3 & 0.3 & 0.3 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

（2）模糊推理

R

$$B' = A' \circ R = \begin{bmatrix} 0.8 \\ 1.0 \\ 0.6 \\ 0.3 \\ 0.0 \end{bmatrix}^T \circ \begin{bmatrix} 0.0 & 0.0 & 0.3 & 0.6 & 1.0 \\ 0.0 & 0.0 & 0.3 & 0.6 & 0.6 \\ 0.0 & 0.0 & 0.3 & 0.3 & 0.3 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

$=(0.0，\ 0.0，\ 0.3，\ 0.6，\ 0.8)$

（3）模糊决策

用最大隶属度法进行决策得风门开度为**5**。
用加权平均判决法和中位数法进行决策得风门开度为**4**。

0.0
0.0
0.3
0.6
0.8

三、状态空间

1.基本概念：

①状态：$Q = [q_1, q_2, \dots, q_n]^T$

②操作：$F = \{f_1, f_2, \dots, f_m\}$

③状态空间：(S, O, S_0, G)

S：状态集合

O：操作算子的集合

S_0：包含问题的初始状态，是 S 的非空子集

G：若干具体状态或满足某些性质的路径信息描述

④求解路径：从 S_0 结点到 G 结点的路径 $S_0 \xrightarrow{O_1} S_1 \xrightarrow{O_2} S_2 \xrightarrow{O_3} \dots \xrightarrow{O_k} G$

⑤状态空间的一个解：一个有限的操作算子序列 $O_1, O_2, \dots, O_k$

四、启发式图搜索策略

1.基本概念

①启发信息：在具体求解中，能够利用与该问题有关的信息来简化搜索过程，称此类信息为启发信息
②启发式搜索：利用启发信息的搜索过程
③估价函数：估算节点“希望”程度的量度
④估价函数值 f(n)：从初始节点经过 n 节点到达目标节点的路径的最小代价估计值 f(n)=g(n)+h(n)

估计 实际 启发

⑤g(n)：从初始节点 S_0 到节点 n 的实际代价，例如搜索的深度
⑥启发函数 h(n)：从节点 n 到目标节点 S_g 的最优路径的估计代价
2.启发函数 h(n)对搜索的影响：
h(n)比重大：降低搜索工作量，但可能导致找不到最优解
~~h(n)比重小：一般导致工作量加大，极限情况下变为盲目搜索~~，但可能可以找到最优解
~~3.A 搜索算法：使用了估价函数 f 的最佳优先搜索~~
估价函数 f(n)=g(n)+h(n)
~~4.A*搜索算法~~
如果某一问题有解，那么利用 A*搜索算法对该问题进行搜索则一定能搜索到解，并且一定能搜索到最优的解而结束。
特性：
①可采纳性：当一个搜索算法在最短路径存在时能保证找到它，则称该算法是可采纳的
当 h(n)=0 时，A*搜索算法变为 BFS
②单调性：若对于所有状态 n_i 和 n_j，有 n_j 是 n_i 的后裔，满足$h(n_i)-h(n_j)\le cost(n_i,n_j)$，且目的状态的启发函数值为 0，则称 h(n)是单调的
A*搜索算法中采用单调性启发函数，可以减少比较代价和调整路径的工作量，从而减少搜索代价
③信息性：在两个 A*启发策略的 h1 和 h2 中，如果对搜索空间中的任一状态 n 都有 $h_1(n)\le h_2(n)$，则 h_2 比 h_1 具有更多的信息性
如果某一搜索策略的 h(n)越大，则 A*算法搜索的信息性越多，所搜索的状态越少
更多的信息性需要更多的计算时间

五、遗传算法

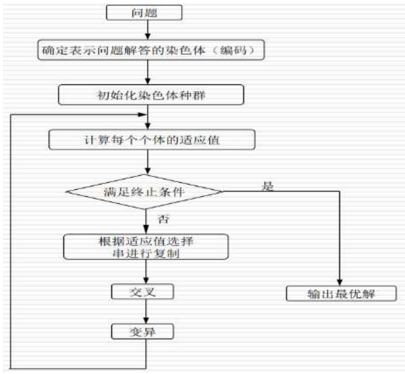
1.遗传算法的基本思想：在求解问题时从多个解开始，然后通过一定的法则进行逐步迭代以产生新的解。遗传算法中每一条染色体，对应着遗传算法的一个解决方案，一般我们用适应性函数来衡量这个解决方案的优劣。所以从一个基因组到其解的适应度形成一个映射。可以把遗传算法的过程看作是一个在多元函数里面求最优解的过程。
2.生物遗传的各个要素在遗传算法中的应用：

生物遗传概念	遗传算法中的应用
适者生存	目标值比较大的解被选择的可能性大
个体	解
染色体	解的编码（字符串、向量等）
基因	解的编码中每一分量
适应性	适应度函数值
群体	根据适应度值选定的一组解（解的个数为群体的规模）
婚配	交叉选择两个染色体进行交叉产生一组新的染色体的过程
变异	编码的某一分量发生变化的过程

3.遗传算法的特点
①对所求解的优化问题没有太多的数学要求；
②利用随机技术指导对一个被编码的参数空间进行高效率搜索；

③采用群体搜索策略，易于并行化；
④仅用适应度函数值来评估个体；
⑤能够非常有效地进行概率意义的全局搜索

4.遗传算法的一般步骤：



开始循环直至找到满意的解：

①评估每条染色体所对应个体适应度。

②遵照适应度越高，选择概率越大的原则从种群中选择两个个体作为父方和母方。

③抽取父母双方的染色体，进行交叉，产生子代。

④对子代的染色体进行变异。

⑤重复②③④步骤，直到新种群的产生。

结束循环。

5.遗传算法的各个步骤

（1）编码

一维染色体编码方法：将问题空间的参数编码为一维排列的染色体的方法

①二进制编码：用若干二进制数表示一个个体，将原问题的解空间映射到位串空间 B={0，1}上，然后在位串空间上进行遗传操作

优点：类似于生物染色体的组成，算法易于用生物遗传理论解释，遗传操作如交叉、变异等易实现；算法处理的模式数最多

缺点：相邻整数的二进制编码可能具有较大的 Hamming 距离，降低了遗传算子的搜索效率；要先给出求解的精度；求解高维优化问题的二进制编码串长，算法的搜索效率低

②实数编码：不必进行数制转换，可直接在解的表现型上进行遗传操作

③多参数级联编码：把每个参数先进行二进制编码得到子串，再把这些子串连成一个完整的染色体；每个子串对应各自的编码参数，因此可以有不同的串长度和参数的取值范围

（2）群体设定

①初始种群的产生：随机产生一定数目的个体，从中挑选最好的个体加到初始群体中，不断迭代直到初始群体中个体数目达到了预先确定的规模

②种群规模的确定：规模太小，遗传算法的优化性能不太好，易陷入局部最优解；规模太大，计算复杂

（3）将目标函数映射成适应度函数

①若目标函数为最大化问题：则$Fit(f(x)) = f(x)$

②若目标函数为最小化问题，则$Fit(f(x)) = \frac{1}{f(x)}$

欺骗问题：在遗传算法中，妨碍适应度值高的个体产生，从而影响遗传算法正常工作的问题

尺度变换：对适应度函数值域的某种映射变换

(i)线性变换：$f' = af + b$满足$f'_{avg} = f_{avg}, f'_{max} = C_{mult} \cdot f_{avg}$

$$a = \frac{f_{avg}}{f_{avg} - f_{min}}, b = \frac{-f_{min} f_{avg}}{f_{avg} - f_{min}}$$

(ii)幂函数变换法：$f' = f^k$

(iii)指数变换法：$f' = e^{-a}$

（4）选择

作用：优胜劣汰，适者生存

个体选择概率分配方法：

①适应度比例方法（蒙特卡罗法）

$$p_{si} = \frac{f_i}{\sum_{i=1}^{M} f_i}$$

②排序方法：

(i)线性排序：群体成员按适应值大小从好到坏排列：$x_1, x_2, \ldots, x_N$，个体 x_i 分配选择概率$p_i = \frac{a-bi}{M(M+1)}$

(ii)非线性排序：将群体成员按适应值从好到坏依次排列，并按下式分配选择概率：

$$p_i = \begin{cases} q(1-q)^{i-1} & i = 1,2,\ldots,M-1 \\ (1-q)^{M-1} & i = M \end{cases}$$

选择个体方法：

①轮盘赌选择：按个体的选择概率产生一个轮盘，轮盘每个区的角度与个体的选择概率成比例。产生一个随机数，它落入转盘的哪个区域就选择相应的个体交叉。

②锦标赛选择方法：从群体中随机选择多个个体，将其中适应度最高的个体保存到下一代。这一过程反复执行，直到保存到下一代的个体数达到预先设定的数量为止。

③随机竞争方法：每次按赌轮选择方法选取一对个体，然后让这两个个体进行竞争，适应度高者获胜。如此反复，直到选满为止。

④最佳个体保存方法：把群体中适应度最高的个体不进行交叉而直接复制到下一代中，保证遗传算法终止时得到的最后结果一定是历代出现过的最高适应度的个体。

（5）交叉

作用：保证种群的稳定性，朝着最优解的方向进化

①基本的交叉算子

(i)一点交叉：在个体串中随机设定一个交叉点，实行交叉时，该点前或后的两个个体的部分结构进行互换，并生成两个新的个体

(ii)二点交叉：随机设置两个交叉点，将两个交叉点之间的码串相互交换

②修正的交叉方法：部分匹配交叉 PMX

（6）变异

作用：保证种群的多样性，避免交叉可能产生的局部收敛

①位点变异：群体中的个体码串，随机挑选一个或多个基因座，并对这些基因座的基因值以变异概率作变动

②逆转变异：在个体码串中随机选择两点（逆转点），然后将两点之间的基因值以逆向排序插入到原位置中

③插入变异：在个体码串中随机选择一个码，然后将此码插入随机选择的插入点中间

④互换变异：随机选取染色体的两个基因进行简单互换

⑤移动变异：随机选取一个基因，向左或者向右移动一个随机位数

6.遗传算法的改进算法：①双倍体遗传算法；②双种群遗传算法；③自适应遗传算法

7.一些其他算法

①爬山法（最速上升爬山法）

算法：从搜索空间中随机产生邻近的点，从中选择对应解最优的个体，替换原来的个体，不断重复上述过程

缺点：对于存在很多局部最优点的问题，通过一个简单的迭代找出全局最优解的机会非常渺茫

②模拟退火

让算法从较大的跳跃开始，使到它有足够的“能量”逃离可能“路过”的局部最优解而不至于限制在其中，当它停在全局最优解附近的时候，逐渐的减小跳跃量，以便使其“落脚 ”到全局最优解上

【例题 1 • 完整解题过程】利用遗传算法求解区间[0,31]上的二次函数 $y=x^2$的最大值

分析：原问题可以转化为在区间[0,31]中搜索能使 y 取最大值的点 a 的问题。那么[0,31]中的点 x 就是个体，函数值 f(x)可以作为 x 的适应度，区间[0,31]就是一个解空间。这样，只需要给出个体 x 的适当染色体编码，该问题就可以用遗传算法来解决。

解：（1）设定种群规模、编码染色体，产生初始种群

将种群规模设定为 4，用 5 位二进制数编码染色体；取下列个体组成初始种群：
S1：s1=13(01101)，s2=24(11000)，s3=8(01000)，s4=19(10011)
（2）定义适应度函数，取适应度函数 f(x)=x*x
（3）计算各代种群中的各个体的适应度，并对染色体进行遗传操作，直到适应度最高的个体，即 31(11111)出现为止
首先计算 S1 中各个体的实用度 f(si)：f(s1)=f(13)=13*13=169，f(s2)=f(24)=24*24=576，f(s3)=f(8)=8*8=64；f(s4)=f(19)=19*19=361；再计算种群 S1 中各个体的选择概率，选择概率的计算公式为：

$$P(x_i) = \frac{f(x_i)}{\sum_{j=1}^{N} f(x_j)}$$

由此可求得：P(s1)=P(13)=0.14，P(s2)=P(24)=0.49，P(s3)=P(8)=0.06，P(s4)=P(19)=0.31
（4）选择-复制
轮盘赌选择法：
①在区间[0,1]内产生一个均匀分布的随机数 r
②若 $r \le q_1$，则染色体 x_1 被选中
③若 $q_{k-1} < r \le q_k (2 \le k \le N)$，则染色体 x_k 被选中，其中 q_i 为染色体 x_i 的累计概率，计算公式为：

$$q_i = \sum_{j=1}^{i} P(x_j)$$

设从区间[0,1]中产生 4 个随机数如下：r1=0.450126，r2=0.110347，r3=0.572496，r4=0.98503

染色体	适应度	选择概率	累计概率	选中次数
s1=01101	169	0.14	0.14	1
s2=11000	576	0.49	0.63	2
s3=01000	64	0.06	0.69	0
s4=10011	361	0.31	1.00	1

复制得群体 s1’=11000(24)，s2’=01101(13)，s3’=11000(24)，s4’=10011(19)
（5）交叉
设交叉率 pc=100%，即 S1 中的全部染色体都参与交叉运算
设 s1’与 s2’配对，s3’与 s4’配对，分别交换后两位基因，得新染色体：
s1’’=11001(25)，s2’’=01100(12)，s3’’=11011(27)，s4’’=10000(16)
（6）变异
设变异率 pm=0.001，群体 S1 中共有 5×4×0.001=0.02 位基因可以变异，显然不足 1 位，所以本轮一传操作不做变异，得到第二代种群 S2：s1=11001(25)，s2=01100(12)，s3=11011(27)，s4=10000(16)
在迭代多代种群中出现适应度最高的染色体 s1=11111，于是遗传操作终止，将染色体“11111”作为最终结果输出，然后将染色体“11111”解码为表现形式，即得所求的最优解：31
将 31 代入函数 $y=x^2$中，即得原问题的解，即函数 $y=x^2$的最大值为 961
【例题 2·思路】求[0,40]范围内的 $y=(x-10)^2$ 的最小值
（1）编码算法选择为“将 x 转化为 2 进制的串”，串的长度为 5 位，等位基因的值为 0 或 1
（2）计算适应度的方法：先将个体串进行解码,转化为 int 型的 x 值,然后使用 $y=(x-10)^2$ 作为其适应度计算函数（由于是最小值，所以结果越小，适应度越好）
（3）正式开始，先设置群体大小为 4，然后初始化群体（在[0,40]范围内随机选取 4 个正整数，编码）
（4）计算适应度 fi，由于是最小值，$fi = \frac{1}{y} = \frac{1}{(x-10)^2}$
（5）计算每个个体的选择概率，选择概率$P = \frac{fi}{\sum fi}$

（6）选择：根据所有个体的选择概率进行淘汰选择，先按照每个个体的选择概率创建一个赌轮，然后选取 4 次，每次先产生一个 0-1 的随机小数，然后判断该随机数落在那个段内就选取相对应的个体，这个过程中，选取概率 P 高的个体将可能被多次选择，而概率低的就可能被淘汰

六、神经网络

1.激活函数

- （1）各种激活函数

①非线性激活函数

(i)硬极限函数（阶跃函数）：$y=\begin{cases}1 & x\geq 0\\ 0 & x<0\end{cases}$

(ii)对称硬极限函数：$y=\begin{cases}1 & x\geq 0\\ -1 & x<0\end{cases}$

②非线性激活函数

(i)Sigmoid 函数（对数-S 形函数，S 型函数）：$y=\frac{1}{1+e^{-\alpha x}}$，α=1

(ii)tanh 函数（双曲正切 S 形函数）：$y=\frac{e^{\alpha x}-e^{-\alpha x}}{e^{\alpha x}+e^{-\alpha x}}$，α=1

Sigmoid 函数的缺点：在输入的绝对值大于某个阈值后，过快进入饱和状态，不再有显著变化，出现梯度消失现象，在实际模型训练中会导致模型收敛缓慢，性能不够理想

ReLU 函数：$f(x)=\begin{cases}0 & x<0\\ x & x\geq 0\end{cases}$

优点：当输入为正时，不存在梯度饱和问题；ReLU 函数中只存在线性关系，因此计算速度快得多

（2）为什么引入激活函数

线性模型表达力不够，激活函数可增加非线性因素；为了增强网络的表达能力，需要激活函数来将线性函数转化为非线性函数

（3）激活函数的性质

①非线性：线性激活层对于深层神经网络没有作用，仍为线性变换；

②连续可微：梯度下降法要求满足连续可微；

③范围最好不饱和：有饱和区间段时，系统优化时梯度似为 0，网络的学习会停止；

④单调性：当激活函数单调时，单层神经网络的误差函数为凸，易优化；

⑤原点处近似线性：当权值初始化为接近 0 的随机值时，网络学习较快

2.单层感知器

①1943 年，麦克洛奇和皮兹提出 M-P 模型

②单层感知器是最简单的一种人工神经网络结构，包含输入层和输出层；输入层只负责接受外部信息，每个输入节点接收一个输入信号；输出层（处理层）具有信息处理能力以及向外部输出处理信息

③局限性：无泛化能力；结构简单，激活函数只能是符号函数；只对线性可分问题收敛；如果存在离群点，则需要花费较多的训练时间

④感知器（单层神经网络，神经元）：是组成神经网络的最小单位，掀起了人工神经网络第一次高潮

⑤存在的问题：单层神经网络无法解决不可线性分割的问题（如：异或门电路）；利用当时最先进的计算机也没有足够计算力完成多层感知器训练所需的超大计算量

3.生物神经元的数学模型

$y_i(t)$：第 i 个神经元的输出；θ_i：第 i 个神经元的阈值；$u_k(t)$：外部输入；a_{ij}、b_{ik}：权值

加权求和：

$$v_i(t)=\sum_{j=1}^{N}a_{ij}y_j(t)+\sum_{k=1}^{M}b_{ik}u_k(t)-\theta_i$$

矩阵形式：$V(t)=AY(t)+BU(t)-\theta$

$A=\{a_{ij}\}_{N\times N}\quad B=\{b_{ik}\}_{N\times M}\quad V=[v_1,\dots,v_N]^T\quad U=[u_1,\dots,u_M]^T\quad \theta=[\theta_1,\dots,\theta_N]^T\quad Y=[y_1,\dots,y_N]^T$

softmax 函数：确保较小的值具有较小的概率，并且不会直接丢弃，函数的分母结合了原始输出值的所有因子，函数获得的各种概率彼此相关；但在零点不可微，且负输入的梯度为零，这意味着对于该区域的激活，权重不会在反向传播期间更新，因此会产生永不激活的死亡神经元

$$softmax:\frac{e^{z_i}}{\sum_{j=1}^{K}e^{z_j}}$$

4.人工神经网络的结构

①前馈型（前向型）

每个神经元只与前一层的神经元相连；当前层只接收前一层的输出；当前层自身的输出只能输出给下一层；各层之间没有反馈；

优点：通常适用于数值数据的监督学习

缺点：不能与顺序数据一起使用

②反馈型（递归网络、回归网络）：每个神经元同时将自身输出作为输入信号反馈给其他神经元，网络的连接图中具有回路，需要工作一段时间后才能稳定

5.人工神经网络的工作方式

①同步（并行）方式：任一时刻神经网络中所有神经元同时调整状态

②异步（串行）方式：任一时刻只有一个神经元调整状态，而其它神经元的状态保持不变

6.人工神经网络的学习

神经网络方法是一种知识表示方法和推理方法，神经网络的学习是指调整神经网络的连接权值或者结构，使输入输出具有需要的特性

7.BP 学习算法（掀起人工神经网络的第二次高潮）

（一）提出

①1974 年，保罗·沃波斯证明了在感知器神经网络中再多加一层，并利用误差的反向传播(BP)来训练人工神经网络，可以解决异或问题

②1985 年，辛顿和鲁梅尔哈特等重新设计了 BP 学习算法，并于 1986 年发表经典论文

• （二）工作原理

基本思想：学习过程由信号的正向传播和误差的反向传播两个过程组成；

正向传播时，把样本的特征从输入层进行输入，信号经过各个隐藏层的处理后，最后从输出层传出；对于网络的实际的输出与期望输出之间的误差，把误差信号从最后一层逐层反传，从而获得各个层的误差学习信号，然后再根据误差学习信号来修正各层神经元的权值；

信号正向传播与误差反向传播，各层调整权值的过程是周而复始地进行；权值不断调整的过程，就是网络学习训练的过程。进行此过程直到网络输出误差减小到预先设置的阈值以下，或者超过预先设置的最大训练次数

（三）工作过程

①第一阶段（网络训练阶段）：对网络的连接权进行学习和调整，以使该网络实现给定样本的输入输出映射关系

②第二阶段（工作阶段）：把实验数据或实际数据输入到网络，网络在误差范围内预测计算出结果

（四）两个问题

①是否存在一个 BP 神经网络能够逼近给定的样本或者函数

BP 定理：给定任意ε>0，对于任意的连续函数，存在一个三层前向神经网络，它可以在任意ε平方误差精度内逼近连续函数

②如何调整 BP 神经网络的连接权，使网络的输入与输出与给定的样本相同

1986 年，鲁梅尔哈特提出 BP 学习算法

(i)神经网络主要是由三个部分组成的：网络架构（输入层、隐藏层、输出层）、激活函数、找出最优权重值的参数学习算法

(ii)BP 算法就是目前使用广泛的一种参数学习算法

(iii)无法直接得到隐层的权值，先通过输出层得到输出结果和期望输出的误差来间接调整隐层的权值

• （五）BP 算法的基本思想

学习过程由信号的正向传播（求损失）与误差的反向传播（误差回传）两个过程组成

1.目标函数（损失函数）：

$$J=\frac{1}{2}\sum_{j=1}^{p_m}\left(y_j^m-d_j\right)^2$$

$J=\frac{1}{2}\sum_{j=1}^{p_m}(y_j^m-d_j)^2$

2.约束函数：

$$u_i^k=\sum_j w_{ij}^{k-1}y_j^{k-1}\qquad i=1,2,\dots,p_k$$

$\cancel{w}z \to a$

$$y_i^k=f_k\left(u_i^k\right)\qquad k=1,2,\dots,m$$

3.连接权值的修正量：

$\cancel{a}z_2 \to y$

$$\Delta w_{ij}^{k-1}=-\varepsilon\frac{\partial J}{\partial w_{ij}^{k-1}}\qquad j=1,2,\dots,p_{k-1}$$

其中，ε是步长，学习率范围为(0,1)，反向梯度

4.学习算法：

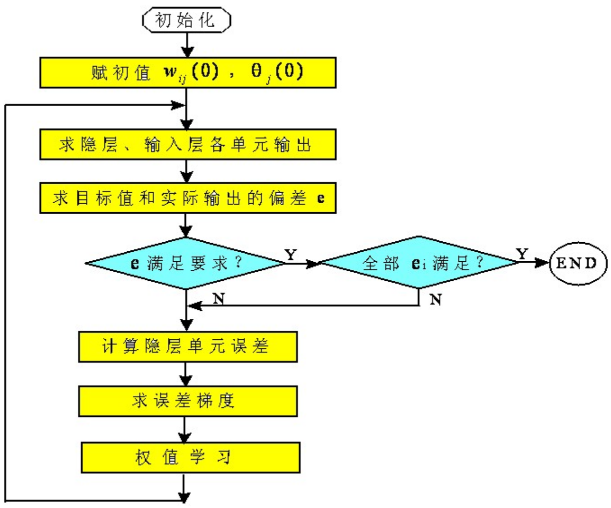
$\Delta w=-\varepsilon da$

当$y_i^k=\frac{1}{1+e^{-u_i^k}}$时，$\Delta w_{ij}^{k-1}=-\varepsilon d_i^k y_j^{k-1}$

$d_i^m=y_i^m(1-y_i^m)(y_i^m-y_i)$——输出层连接权调整公式

$d_i^k=y_i^k(1-y_i^k)\sum_{l=1}^{p_{k+1}}w_{li}^{k+1}d_l^{k+1}$——隐藏层连接权调整公式

$d=y(1-y)(y-y_i)$



• （五）BP 学习算法应用

1.常用模式

BP 神经网络模型，采用非线性激活函数 Sigmoid 函数，具有输入层、隐藏层、输出层三个层次

输入层：负责接收外界刺激（外部数据）

隐藏层：负责增加计算能力，以解决困难问题，但隐藏层神经元不能无限增加，否则会出现过拟合现象

输出层（决策层）：负责进行决策

一般来说，当数据和决策问题确定之后，输入层和输出层的节点个数是固定的，唯一可变的是隐藏层；层与层之间是通过神经键（权重）完成连接；

训练好之后的 BP 神经网路模型是前馈的，但在进行模型训练的时候是从输出层到输入层，是反馈的

2.构建 BP 神经网络

①只有隐藏层和输出层的神经元才有构造，且每个神经元的构造都是一模一样的，而输入层的神经元只负责接收外来数据，数据一旦输入，便会立即输出，故不具有构造

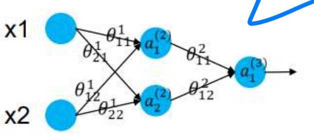
②神经元由上一层神经元与本层神经元交叉形成的神经键、权重加总、常数项、激活函数组成

③$w_{0j}, w_{1j}, \ldots, w_{nj}$ 表示上一层神经元对于该神经元的影响权重，可正可负，范围为[0,1]；常数项表示神经元自身的偏好；激活函数 f 表示一个以权重加总与常数项之和为自变量的函数，决定神经元的输出结果

④在 BP 神经网络中，一般采用 S 型函数作为激活函数，表达式为$\frac{1}{1+e^{-x_j}}$，其中 x_j 是权重加总与常数项之和，称为组合函数的值；当组合函数的值为负时，激活函数的值小于 0.5；否则激活函数的值大于 0.5

【例题 1】

构建一个BP神经网络（网络结构如下图）完成表中的数据分类，设初始连接权重全部为0.5，学习率为0.1，输入层到隐含层，隐含层到输出层的激活函数为Sigmoid函数，代价函数为均方误差（即Cost=$\frac{1}{2}\left(a_1^{(3)}-y\right)^2$），要求误差e为0.01，当输入样本（1，0，1）时，计算BP算法执行第一轮后各连接权重对应的值。其中$a_j^{(l)}$表示第L层第j个神经元的激活值，θ_{ji}^{l}表示第l层的第i个神经元与第l+1层的第j个神经元的连接权重。



x1	x2	y
1	1	0
1	0	1
0	1	1
0	0	0

①前向传播：

$$a_1^{(2)}=g(\theta_{11}^{1}x_1+\theta_{12}^{1}x_2)=g(0.5*1+0.5*0)=g(0.5)=Sigmoid(0.5)=0.6225$$

$$a_2^{(2)}=g(\theta_{21}^{1}x_1+\theta_{22}^{1}x_2)=g(0.5*1+0.5*0)=g(0.5)=Sigmoid(0.5)=0.6225$$

$$a_1^{(3)}=g\left(\theta_{11}^{2}a_1^{(2)}+\theta_{12}^{2}a_2^{(2)}\right)=g(0.5*0.6225+0.5*0.6225)=g(0.6225)=Sigmoid(0.6225)=0.6508$$

②误差反向传播：

$$\delta_1^{(3)}=a_1^{(3)}-y=0.6508-1=-0.3492$$

$$\delta_1^{(2)}=\theta_{11}^{2}\delta_1^{(3)}g'(z)=\theta_{11}^{2}\delta_1^{(3)}a_1^{(2)}\cdot\left(1-a_1^{(2)}\right)=-0.041$$

$$\delta_2^{(2)}=\theta_{12}^{2}\delta_1^{(3)}g'(z)=\theta_{12}^{2}\delta_1^{(3)}a_2^{(2)}\cdot\left(1-a_2^{(2)}\right)=-0.041$$

注：Sigmoid 函数求导——$g'(z)=g(z)\cdot(1-g(z))$，$y=\frac{1}{2}x^2$求导结果为 x

③偏导计算

$$\frac{\partial J(\theta)}{\partial\theta_{11}^{1}}=x_1\cdot\delta_1^{(2)}=1*(-0.041)=-0.041$$

$$\frac{\partial J(\theta)}{\partial\theta_{12}^{1}}=x_2\cdot\delta_1^{(2)}=0*(-0.041)=0$$

$$\frac{\partial J(\theta)}{\partial \theta_{21}^{1}} = x_1 \cdot \delta_2^{(2)} = 1 * (-0.041) = -0.041$$

$$\frac{\partial J(\theta)}{\partial \theta_{22}^{1}} = x_2 \cdot \delta_2^{(2)} = 0 * (-0.041) = 0$$

$$\frac{\partial J(\theta)}{\partial \theta_{11}^{2}} = a_1^{(2)} \cdot \delta_1^{(3)} = 0.6225 * (-0.3492) = -0.2174$$

$$\frac{\partial J(\theta)}{\partial \theta_{12}^{2}} = a_2^{(2)} \cdot \delta_1^{(3)} = 0.6225 * (-0.3492) = -0.2174$$

注：$\frac{\partial J(\theta)}{\partial \theta_{ij}^{l}} = a_j^{(l)} \delta_i^{(l+1)}$

④梯度下降

$$\theta_{11}^{1} = \theta_{11}^{1} - \alpha \frac{\partial J(\theta)}{\partial \theta_{11}^{1}} = 0.5 - 0.1 * (-0.041) = 0.5041$$

$$\theta_{12}^{1} = \theta_{12}^{1} - \alpha \frac{\partial J(\theta)}{\partial \theta_{12}^{1}} = 0.5 - 0.1 * 0 = 0.5$$

$$\theta_{21}^{1} = \theta_{21}^{1} - \alpha \frac{\partial J(\theta)}{\partial \theta_{21}^{1}} = 0.5 - 0.1 * (-0.041) = 0.5041$$

$$\theta_{22}^{1} = \theta_{22}^{1} - \alpha \frac{\partial J(\theta)}{\partial \theta_{22}^{1}} = 0.5 - 0.1 * 0 = 0.5$$

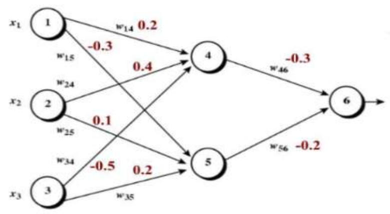
$$\theta_{11}^{2} = \theta_{11}^{2} - \alpha \frac{\partial J(\theta)}{\partial \theta_{11}^{2}} = 0.5 - 0.1 * (-0.2174) = 0.52174$$

$$\theta_{12}^{2} = \theta_{12}^{2} - \alpha \frac{\partial J(\theta)}{\partial \theta_{12}^{2}} = 0.5 - 0.1 * (-0.2174) = 0.52174$$

【例题 2】假设训练数据(x1,x2,x3)为(1,0,1)，各个神经键的权重值以及隐藏层和输出层神经元的常数项随机取值如下：

W_{14}	W_{15}	W_{24}	W_{25}	W_{34}	W_{35}	W_{46}	W_{56}	θ_4	θ_5	θ_6
0.2	-0.3	0.4	0.1	-0.5	0.2	-0.3	-0.2	-0.4	0.2	0.1

BP 神经网络结构图如下：



（1）以隐藏层 4 号节点为例，其输出值计算如下：1*0.2+0*0.4+1*(-0.5)+(-0.4)=-0.7　　$1/(1+e^{0.7})$=0.332
同理可得，4、5、6 号节点的输出值依次为 0.332、0.525、0.474

（2）假设实际的结果应当为 1，但采用随机赋予权重值的方法得到输出值为 0.474，与目标值 1 差距过大，因而要对神经网络进行修正

（3）修正权重和常数项值需要一个标准：平方误差；平方误差=$\sum_j (T_j - O_j)^2$，其中 T_j 为训练数据实际值，O_j 为神经网络输出值；若误差确定为 0.01，则到达到误差要求，停止循环

•（4）新的权重值=旧的权重值+学习效率*前节点的输出值*后节点的误差值

假设学习率为 0.1，则 4 号节点的新权重=(-0.3)+0.1*0.332*(1-0.474)=-0.28

其他知识（概念性的知识）

• 一、人工智能的产生与发展

（一）产生

公元前：亚里士多德提出的三段论仍是演绎推理的基本依据；

培根：系统地提出归纳法

莱布尼茨：万能符号、推理计算

布尔：用符号语言描述思维活动的基本推理法则

图灵：1936 年，图灵机

阿塔纳索夫：1937-1941 年间，世界上第一台电子计算机：阿塔纳索夫-贝瑞计算机（ABC）

麦克洛奇与匹兹：1943 年，第一个神经网络模型（M-P 模型）

图灵测试：1950 年

麦卡锡（人工智能之父）：1956 年**达特茅斯会议**提出“人工智能”这一术语

（二）发展

1956-1970：人工智能的形成期（1969 国际人工智能联合会议，1970 创刊了国际性的人工智能杂志）

1970-2010：几起几落的曲折发展期（1966 美国顾问委员会裁定没有很近的实现前景，1977 费根鲍姆在第五届国际人工智能联合会上提出“知识工程”概念，1986 年以后：集成发展时期）

2011-：大数据驱动的飞速发展期（2006：Hinton 提出深度学习方法，2017.10：Hinton 提出胶囊网络；2019 提出了堆叠胶囊自动编码器(SCAE)）

人工智能大体可分为专用人工智能和通用人工智能；目前主要是面向特定任务的专用人工智能，通用人工智能尚处于起步阶段；人工智能的发展方向应该是从专用智能向通用智能。

• 二、人工智能的主要应用领域

自动定理证明、博弈、模式识别、机器视觉、自然语言理解、机器翻译、智能信息检索、专家系统、自动程序设计、机器人、无人驾驶、组合优化问题、智慧物流、人工神经网络、分布式人工智能与多智能体、智能仿真、智能 CAD、智能 CAI、智能管理与智能决策、智能多媒体系统、人工生命、智能操作系统、智能计算机系统、智能通信、智能网络系统、云端人工智能

三、第一章其他琐碎知识点

（一）20 世纪三大科学技术成就：空间技术、原子能技术、人工智能

（二）自然界的四大奥秘：物质的本质、宇宙的起源、生命的本质、智能的发生

（三）智能是知识与智力的总和，知识是一切智能行为的基础，智力是获取知识并应用知识求解问题的能力

（四）智能的特征：感知能力、记忆与思维能力、学习能力、行为能力（表达能力）

思维：逻辑思维（抽象思维）、形象思维（直感思维）、顿悟思维（灵感思维）

四、知识

（一）知识是在长期的生活及社会实践中、在科学研究及实验中积累起来的对客观世界的认识与经验

（二）知识的相对正确性：任何知识都是在一定的条件及环境下产生的，在这种条件及环境下才是正确的

（三）知识的不确定性：随机性引起的不确定性、模糊性引起的不确定性、经验性引起的不确定性、不完全性引起的不确定性

• 五、推理

（一）定义：由一个或几个已知事实（证据）根据某种策略推出结论的过程

（二）分类：
1.演绎推理：一般→个别，三段论式
2.归纳推理：个别→一般，完全归纳推理（必然性推理）、不完全归纳推理（非必然性推理）
3.默认推理（缺省推理）：知识不完全的情况下假设某些条件已经具备所进行的推理
（三）方向：
1.正向推理：简单易实现，但目的性不强，效率低
2.逆向推理（目标驱动推理）：以某个假设目标作为出发点，目的性强，但起始目标的选择有盲目性
3.混合推理：先正向后逆向、先逆向后正向

• 六、搜索
（一）概念
1.盲目搜索：在不具有对特定问题的任何有关信息的条件下，按固定的步骤（依次或随机调用操作算子）进行的搜索
2.启发式搜索：考虑特定问题领域可应用的知识，动态地确定调用操作算子的步骤，优先选择较适合的操作算子，尽量减少不必要的搜索，以求尽快地到达结束状态
（二）分类
1.数据驱动：从初始状态出发的正向搜索
2.目的驱动：从目的状态出发的逆向搜索
3.双向搜索：从开始状态出发作正向搜索，同时又从目的状态出发作逆向搜索，直到两条路径在中间的某处汇合为止

七、遗传算法相关概念
1.智能优化方法通常包括两大类方法：进化计算、群智能
2.进化算法是基于自然选择和自然遗传等生物进化机制的一种搜索方法，包括遗传算法、遗传规划、进化策略、进化规划等
3.进化计算：是一系列基于生物进化原理（如自然选择和遗传）的问题解决技术的统称，主要用于求解优化问题，试图找到全局最优解；该类算法从随机生成一组潜在解开始，然后通过迭代更新这些可能的解得到一个新的种群；更新是通过迭代应用选择、交叉和变异操作来完成的，这个过程随机丢弃不好的解，并进化出更适合（更好）的解
4.遗传算法：是一类借鉴生物界自然选择和自然遗传机制的随机搜索算法；本质是一种高效、并行、全局搜索的方法，能在搜索过程中自动获取和积累有关搜索空间的知识，并自适应地控制搜索过程以求得最佳解；非常适用于处理传统搜索方法难以解决的复杂和非线性优化问题；可广泛应用于组合优化、机器学习、自适应控制、规划设计和人工生命等领域
5.1975 年，美国 J. Holland 出版了《自然系统和人工系统的适配》；DeJong 完成了重要论文《遗传自适应系统的行为分析》

• 八、神经网络相关概念
1.人工神经网络是一个用大量简单处理单元经广泛连接而组成的人工网络，是对人脑或生物神经网络若干基本特性的抽象和模拟