

遗传算法

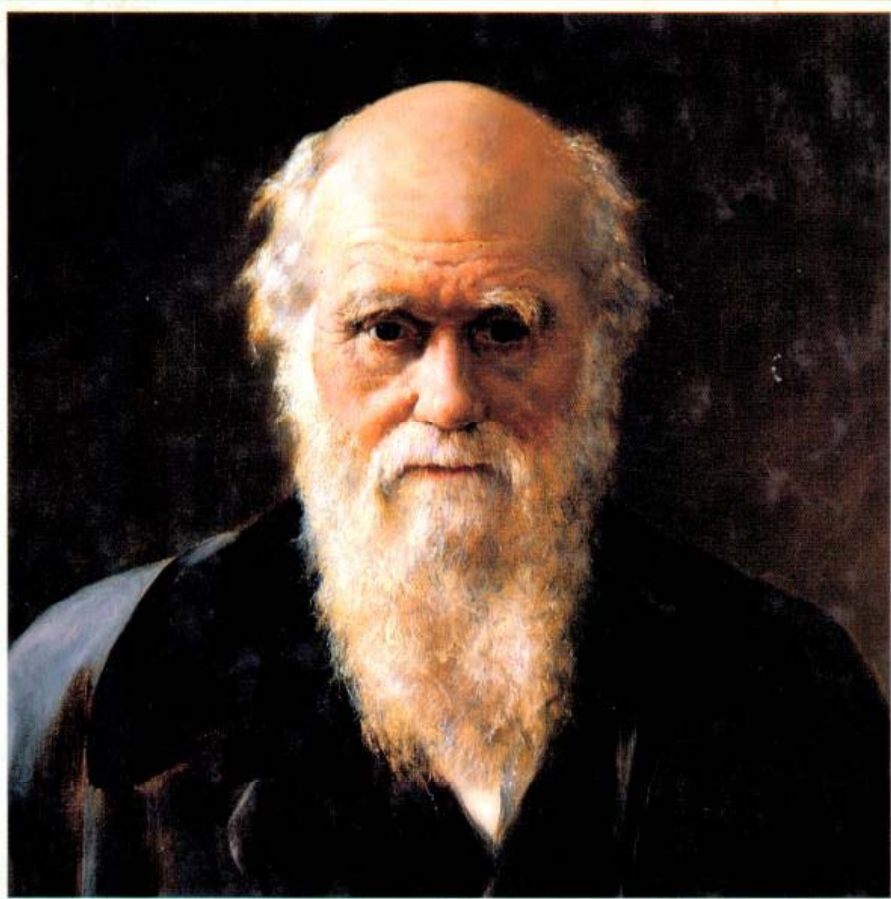
- 群体的力量!



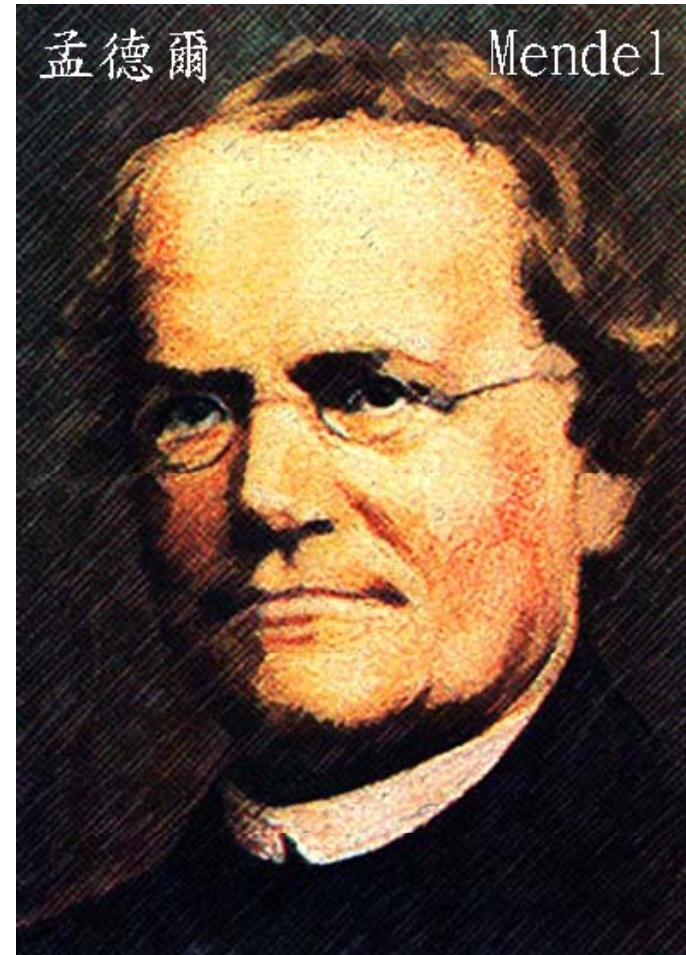
遗传算法

- 达尔文的**进化论**:
 - 自然选择，适者生存
- 孟德尔与摩根的**遗传学理论**:
 - 基因是决定生物特征的最基本的物质单元，基因在染色体上以一定的顺序和结构排列，每个基因有特殊的位置并控制生物的某些特性。
 - 基因组合的特异性决定了生物体的多样性，基因结构的稳定性保证了生物物种的稳定性，而基因的杂交和变异使生物进化成为可能。

遗传算法



Charles Robert Darwin, 1809-1882



Gregor Mendel, 1822-1884

遗传算法

- 生物进化过程的发生需要的**基本条件**:
 - 存在由多个生物个体组成的**种群**;
 - 生物个体之间存在着**差异**, 或群体具有**多样性**;
 - 生物能够自我**繁殖**;
 - 不同个体具有不同的**环境生存能力**, 具有优良基因结构的个体繁殖能力强, 反之则弱。
 - 存在竞争 (**优胜劣汰**) 。

遗传算法

- 遗传算法基本思想
 - 将问题的候选解编码为二进制位串，称为**染色体**（**chromosome**）
 - 生成包含若干不同染色体的**种群**（**population**）
 - 根据每个染色体相对于某种**适应度函数**得到一个**适应度**（**fitness**）
 - 保留适应度较高的染色体，淘汰剩余的染色体
 - “生存”下来的染色体得以**复制**（**replication**）下一代，同时以一定的概率**交叉**（**crossover**）和**变异**（**mutation**），以产生新的个体
 - 每一代均比上一代具有更好的平均性能
 - 通过若干代的进化，找到最优解

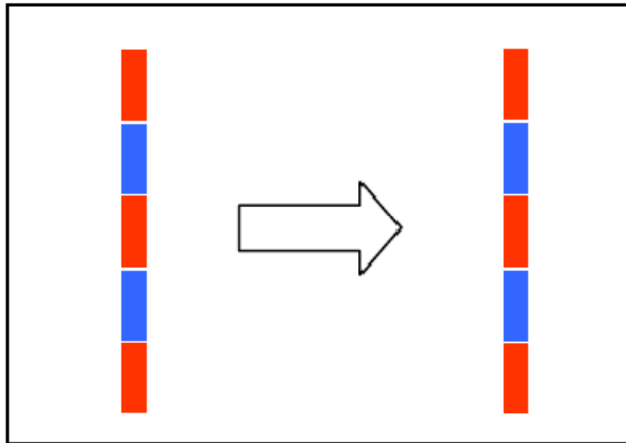
遗传算法

- 基本遗传算法

```
1 begin initialize  $\theta, P_{co}, P_{mut}, L$   $N$ -bit chromosomes
2   do Determine fitness of each chromosome,  $f_i, i = 1, \dots, L$ 
3     Rank the chromosomes
4     do Select two chromosomes with highest score
5       if  $Rand[0, 1) < P_{co}$  then crossover the pair at a randomly chosen bit
6         else change each bit with probability  $P_{mut}$ 
7         Remove the parent chromosomes
8       until  $N$  offspring have been created
9     until Any chromosome's score  $f$  exceeds  $\theta$ 
10    return Highest fitness chromosome (best classifier)
11  end
```

遗传算法

- 三种基本操作
 - 复制 (**replication**)
 - 染色体被原样复制一遍，不发生任何改变



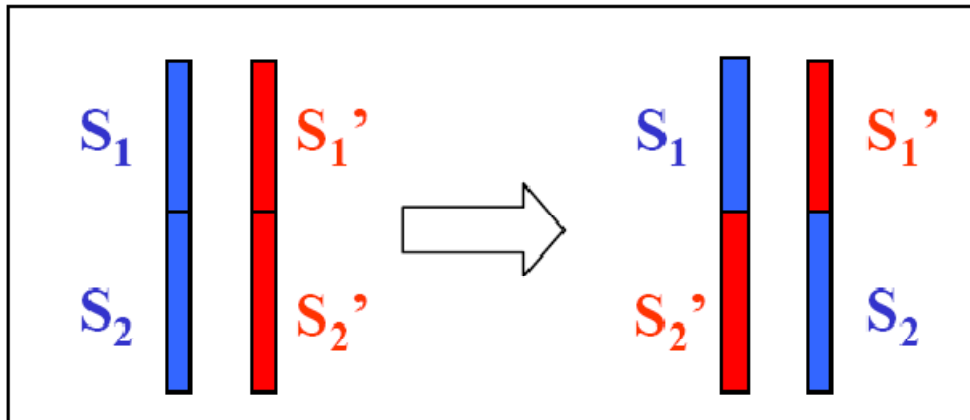
```
11010100101001010101111010100011111010010  
-----  
11010100101001010101111010100011111010010
```

遗传算法

- 三种基本操作

- 交叉 (**crossover**)

- 两条染色体通过交换片段产生两条新的染色体
 - 在染色体上随机确定一个位置并截断
 - A染色体的第一部分与B染色体的第二部分连接
 - A染色体的第二部分与B染色体的第一部分连接



A 11010100101001010101111010100011111010010

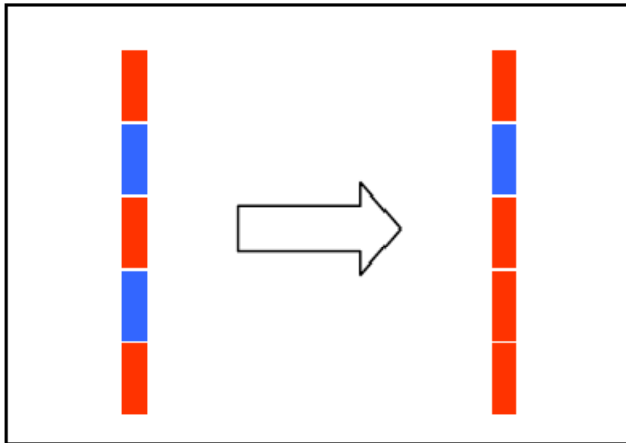
B 00101100001010001010100001010110101001110

11010100101001010101111011010110101001110

00101100001010001010100000100011111010010

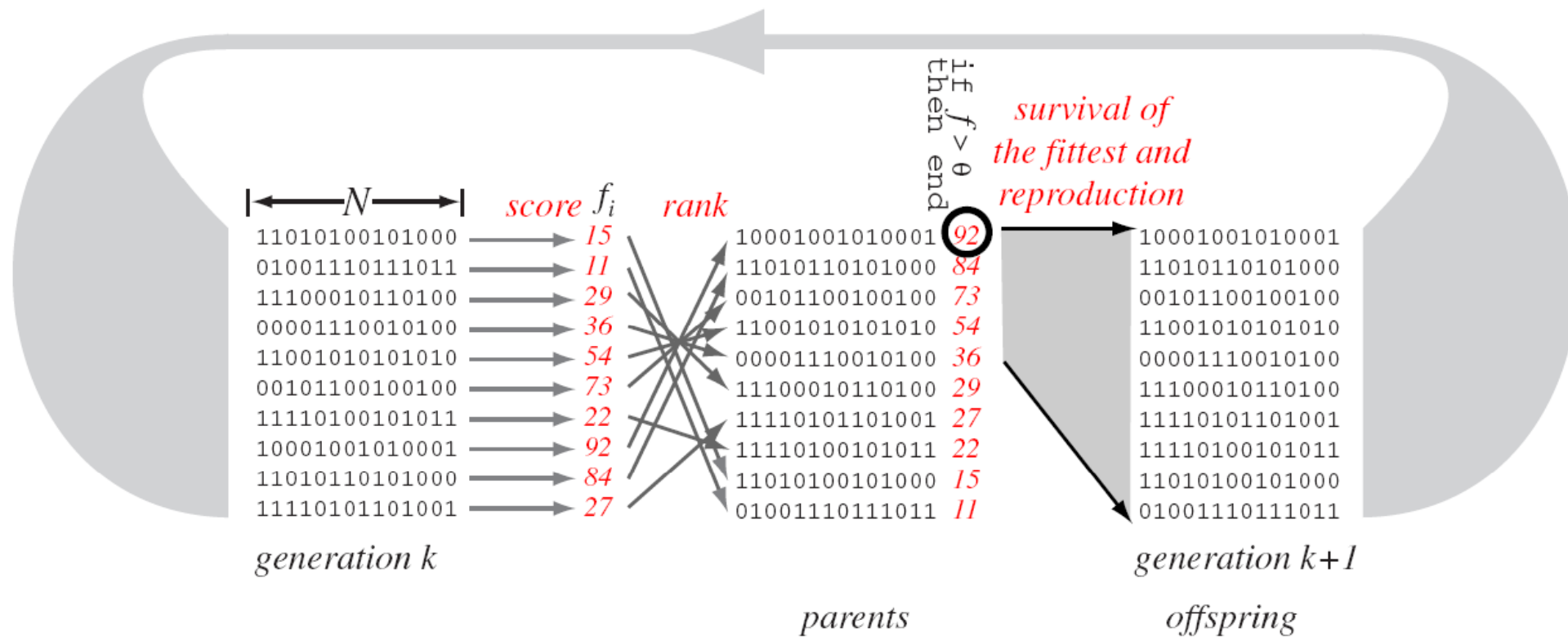
遗传算法

- 三种基本操作
 - 变异 (**mutation**)
 - 染色体中每个位以一个很小的概率改变自身 (0变1, 或1变0)



11010100101001010101111010100011111010010
110111001000111010111110110011101010010

遗传算法



遗传算法

- 染色体编码

如何将一个解编码为二进制位串

- 任意形式的数据都可以表示为二进制串
 - 整数
 - 实数
 - 虚数
 - 向量
 - 矩阵
 - 图像
 - 字符
 -

遗传算法

- 选择

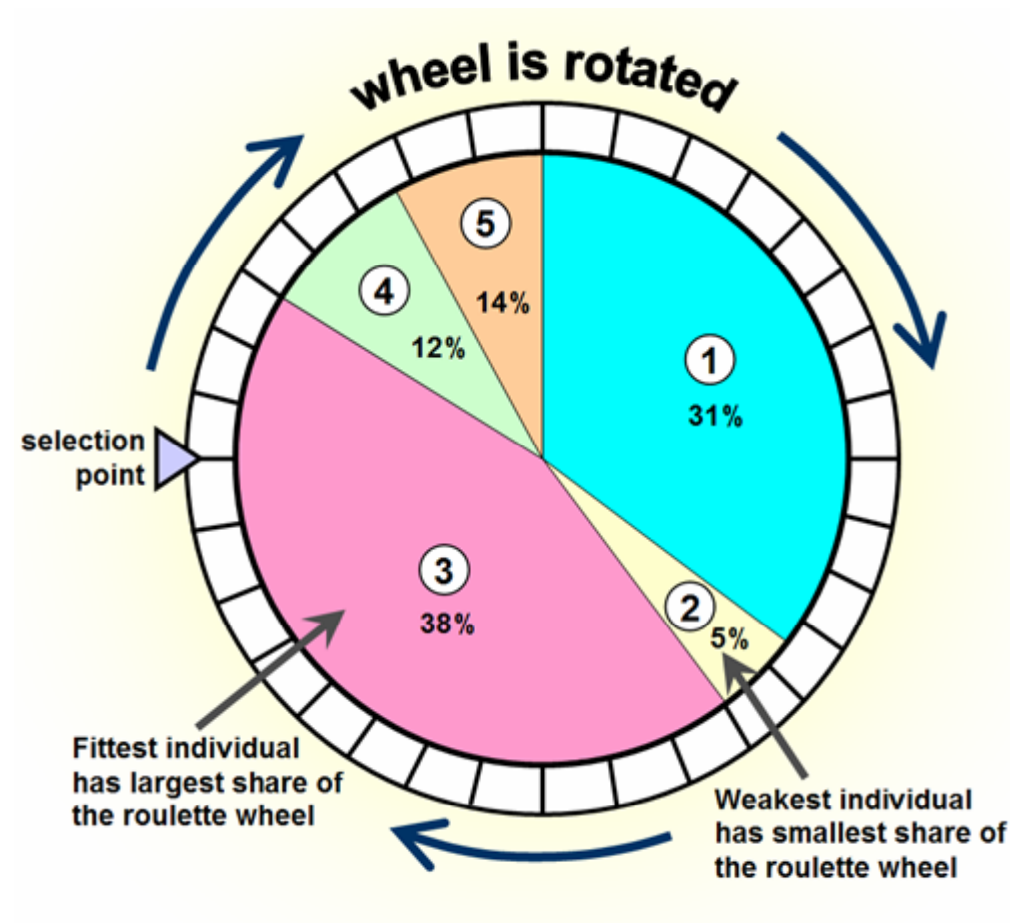
确定某一代中哪些染色体可以为下一代提供遗传信息

- 直接选择适应度最高的若干染色体可能会降低个体间的差异度
- 更好的选择方法是：
 - 以较大的概率选择适应度得分高的染色体
 - 以较小的概率选择适应度得分低的染色体

让适应度分数不高的染色体也有一定机会进入下一代，以保留种群的多样性

遗传算法

- 选择
 - 赌轮选择 (**roulette wheel selection**)



遗传算法

- 例子

用遗传算法求 $f(x) = -\frac{1}{4}x^2 + 2x + 5$ 的在 $x \in [0, 10]$ 范围内的最大值

- 编码

- 假设 x 的精度为小数点后两位
 - 二进制编码为 $x \times 100$ 的二进制表示
 - $x \times 100$ 的范围为 $[0, 1000]$ ，所以需要10位二进制数

遗传算法

- 例子

用遗传算法求 $f(x) = -\frac{1}{4}x^2 + 2x + 5$ 的在 $x \in [0,10]$ 范围内的最大值

- 初始种群生成

No	x	$100x$	染色体
1	1.05	105	0001101001
2	9.62	962	1111000010
3	2.55	255	0011111111
4	9.07	907	1110001011
5	8.87	887	1101110111

遗传算法

- 例子

用遗传算法求 $f(x) = -\frac{1}{4}x^2 + 2x + 5$ 的在 $x \in [0,10]$ 范围内的最大值

- 选择

No	染色体	x	$f(x)$	比例(%)
1	0001101001	1.05	6.82	31
2	1111000010	9.62	1.11	5
3	0011111111	2.55	8.48	38
4	1110001011	9.07	2.57	12
5	1101110111	8.87	3.08	14
总和			22.05	100

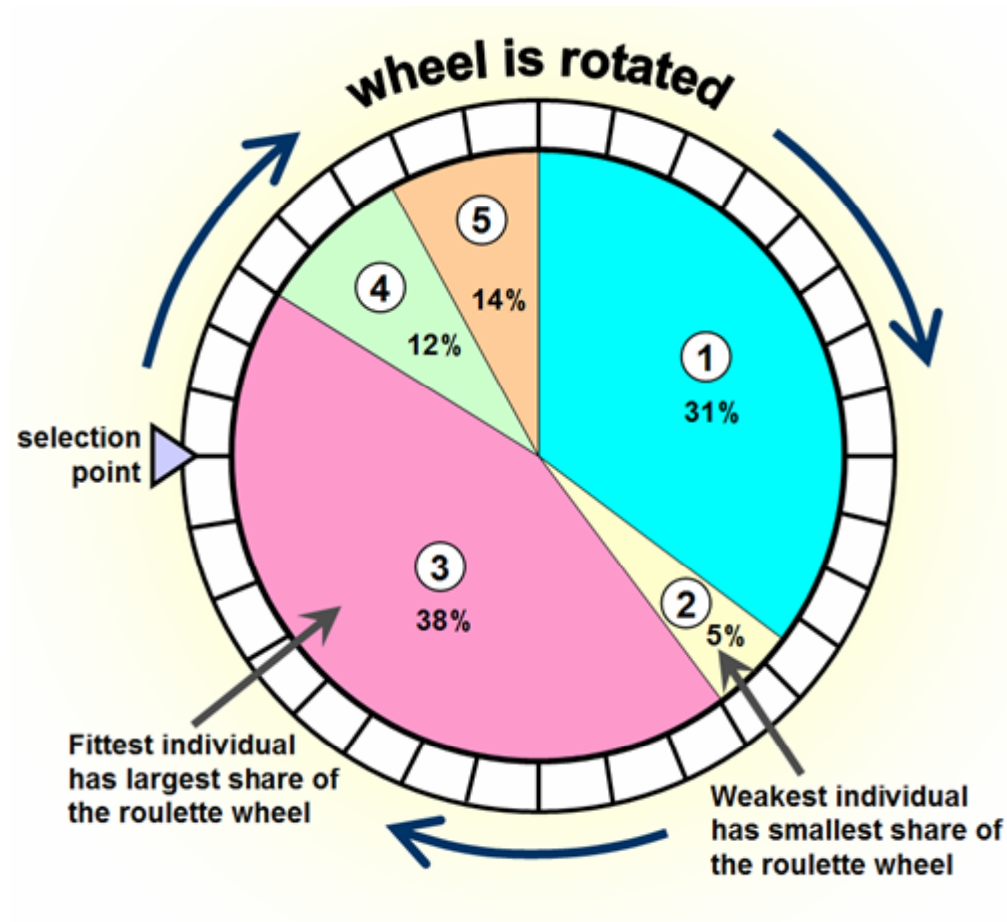
遗传算法

- 例子

用遗传算法求 $f(x) = -\frac{1}{4}x^2 + 2x + 5$ 的在 $x \in [0,10]$ 范围内的最大值

- 选择

赌轮选择



遗传算法

- 例子

用遗传算法求 $f(x) = -\frac{1}{4}x^2 + 2x + 5$ 的在 $x \in [0,10]$ 范围内的最大值

- 交叉

- 以一定的概率 P_{co} 将选择出的两个染色体交叉

x=2.55

0011111111

0001101001

x=1.05



x=2.49

0011111001

0001101111

x=1.11

遗传算法

- 例子

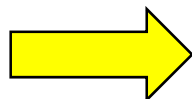
用遗传算法求 $f(x) = -\frac{1}{4}x^2 + 2x + 5$ 的在 $x \in [0,10]$ 范围内的最大值

- 变异

- 以一定的概率 P_{mut} 将选择出的染色体变异

x=2.55

0011111111



x=2.39

0011101111

小结

- 遗传算法
 - 三种基本操作
 - 复制 (**replication**)
 - 交叉 (**crossover**)
 - 变异 (**mutation**)
 - 基本过程
 - 编码
 - 生成种群
 - 计算适应度
 - 选择
 - 交叉
 - 变异