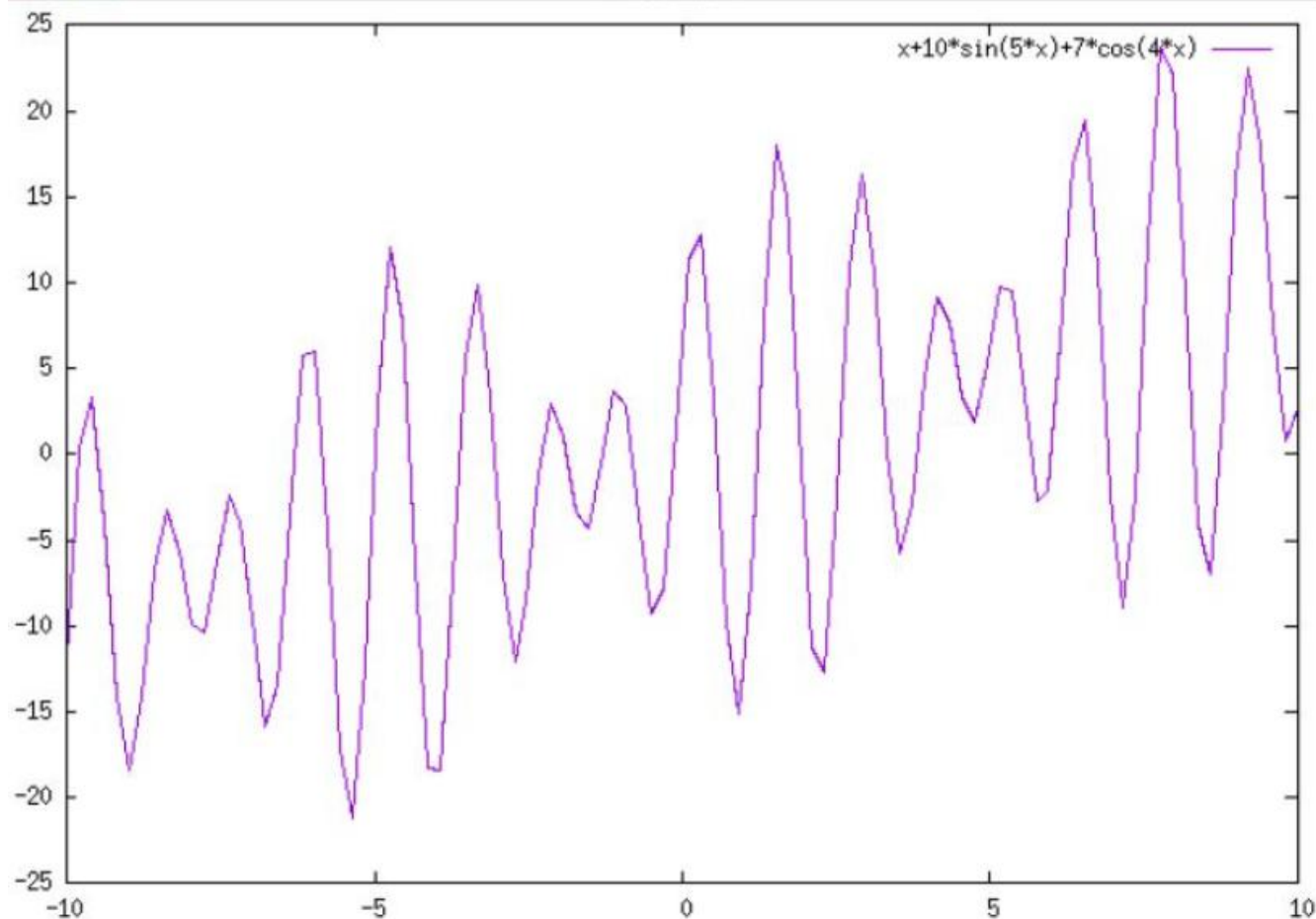

遗传算法

(Genetic Algorithm,GA)

问题背景：

求解函数 $f(x) = x + 10\sin(5x) + 7\cos(4x)$ 在区间 $[0, 9]$ 的最大值。



计算智能

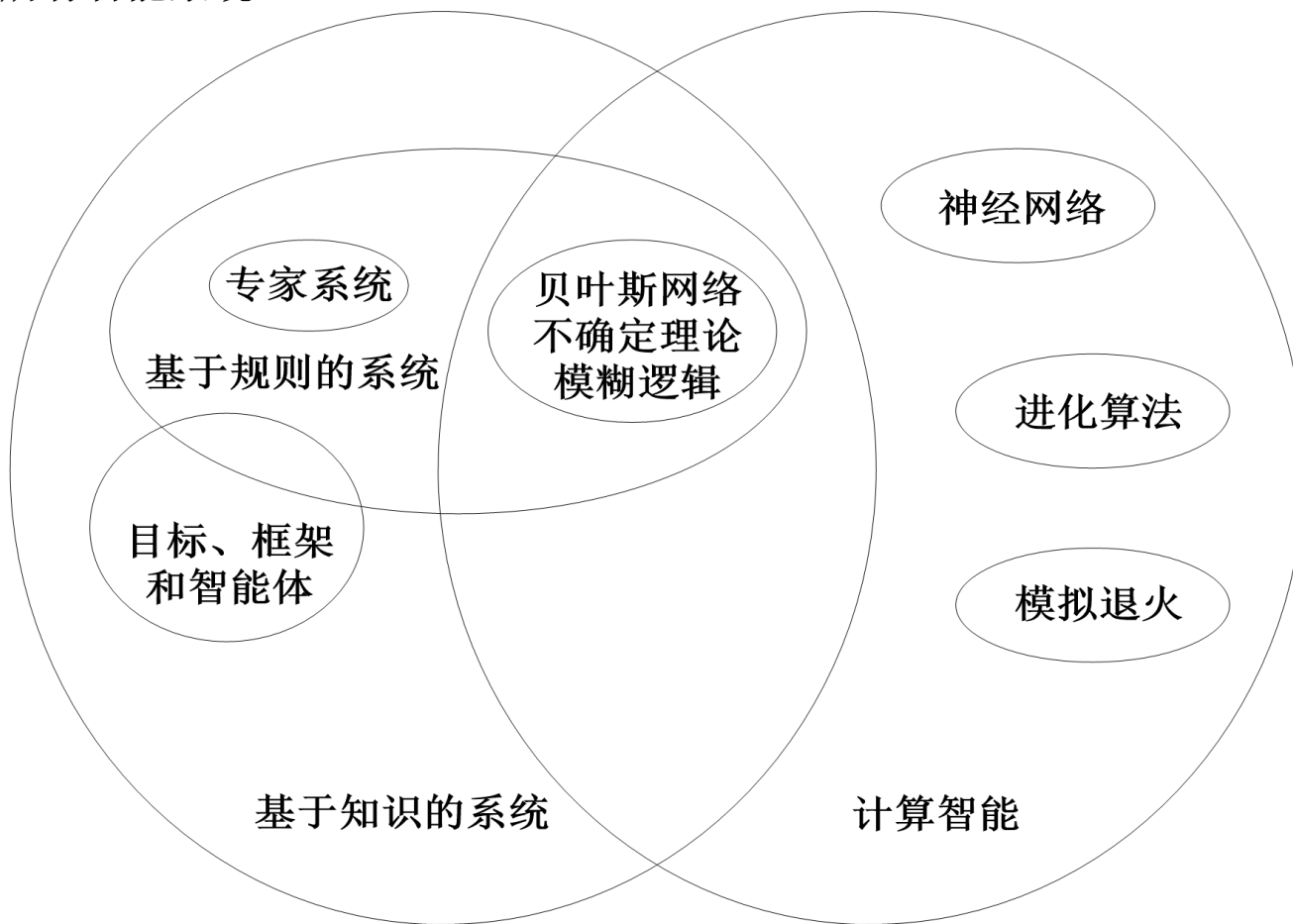
主要包括神经计算、进化计算和模糊计算等。

它们分别从不同的角度模拟人类的智能活动，以使计算机具有智能。

进化计算

遗传算法、免疫算法、进化规划和进化策略的基本思想具有相似之处，这类方法统称为“进化计算” (Evolutionary Computation) 。

所有智能系统



1. 遗传算法起源

- 60年代中期，美国Michigan大学的J. H. Holland教授提出借鉴生物自然遗传的基本原理用于自然和人工系统的自适应行为研究和串编码技术；
- 1967年，他的学生J. D. Bagley在博士论文中首次提出“遗传算法(Genetic Algorithms)”一词；
- **1975年**，Holland出版了著名的“Adaptation in Natural and Artificial Systems”，标志**遗传算法的诞生**。

-
- 70年代初，Holland提出了“模式定理”（Schema Theorem），一般认为是“遗传算法的基本定理”，从而奠定了遗传算法研究的理论基础；
 - 1985年，在美国召开了第一届遗传算法国际会议，并且成立了国际遗传算法学会(ISGA, International Society of Genetic Algorithms)；
 - 1991年，L. Davis编辑出版了《遗传算法手册》，其中包括了遗传算法在工程技术和社会生活中大量的应用实例。

2.1 生物进化理论和遗传学的基本知识

达尔文的自然选择说

- **遗传 (heredity)**：子代和父代具有相同或相似的性状，保证物种的稳定性；
- **变异 (variation)**：子代与父代，子代不同个体之间总有差异，是生命多样性的根源；
- **生存斗争和适者生存**：具有适应性变异的个体被保留，不具适应性变异的个体被淘汰。

遗传学基本概念与术语

- 染色体（chromosome）：遗传物质的载体；
- 脱氧核糖核酸（DNA）：大分子有机聚合物，双螺旋结构；



- 遗传因子（gene）：DNA或RNA长链结构中占有一定位置的基本遗传单位；

达尔文的自然选择学说认为：

- 遗传与变异是决定生物进化的内在因素。

遗传，保持物种的特性，

变异，使物种适应新环境而不断向前发展

- 遗传物质的载体是染色体

染色体是由DNA和蛋白质两种物质组成

- 基因

基因，染色体上具有控制生物性状的DNA片段

一条染色体只有一个DNA分子组成，

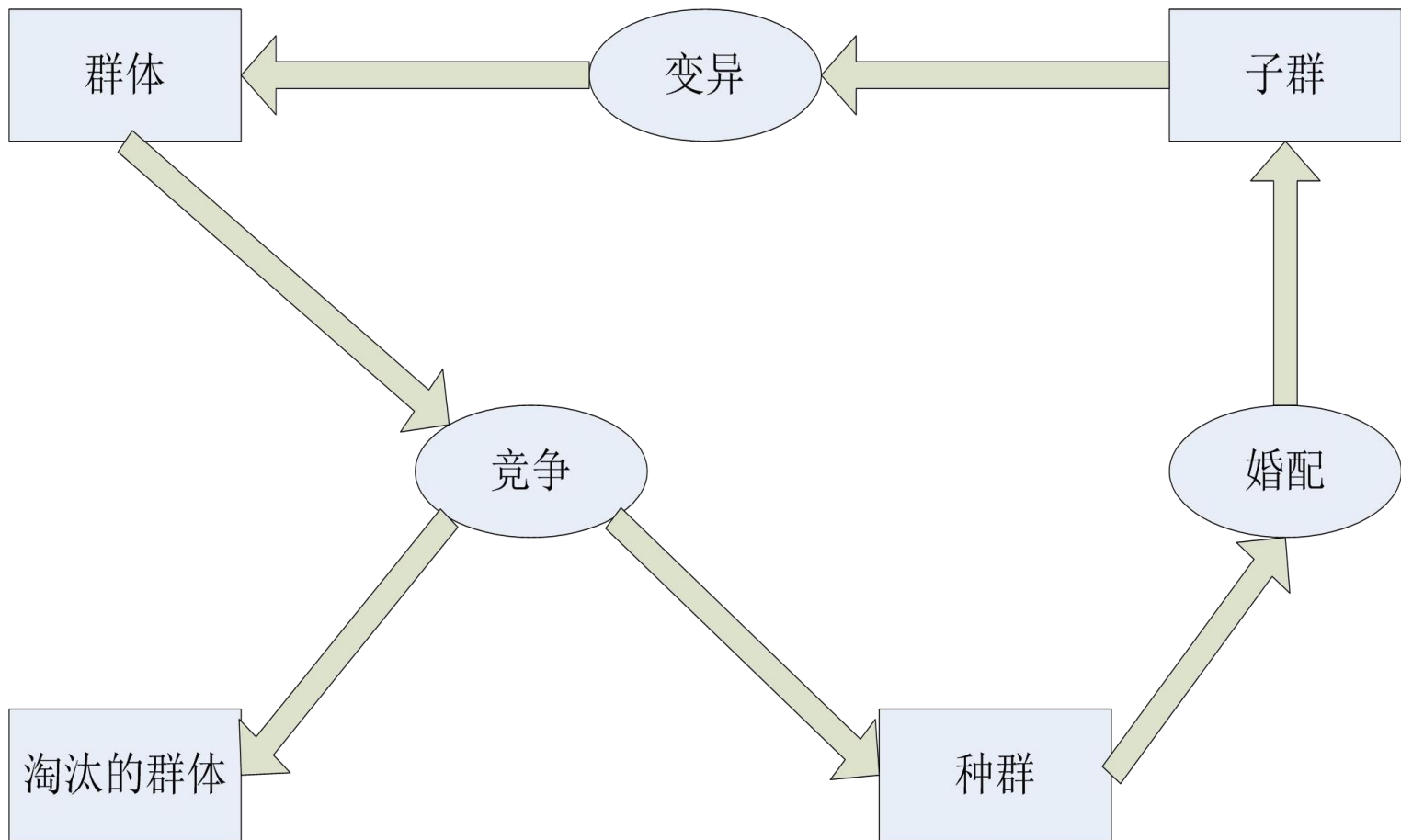
一个DNA子上有许多个基因

- 基因作用

基因储存着遗传信息，可以准确复制，也能够发生突变

2.2 生物遗传与进化的规律

- 染色体由基因构成，决定了生物的形状
- 生物所有遗传信息包含在染色体中，
- 同源染色体的交叉或变异生成新的物种，使生物呈现新的性状
- 适应能力强的基因或染色体，有更多的机会遗传到 下一代



生物进化循环图

3 遗传算法

- ①该算法是根据大自然中**生物体进化规律**设计提出的。
- ②模拟达尔文自然选择和遗传学机理的生物进化过程的计算模型，是一种**通过模拟自然进化过程搜索最优解的方法**
- ③该算法通过数学的方式,利用计算机仿真运算,将问题的求解过程转换成类似生物进化中的染色体基因的交叉、变异等过程。
- ④在求解较为复杂的组合优化问题时,相对一些常规的优化算法,通常能够较快地获得较好的优化结果

遗传算法的应用领域

● 函数优化

非线性、多模型、多目标的函数优化问题，采用其他优化方法较难求解，而遗传算法却可以得到较好的结果

● 自动控制

控制器参数的优化

神经网络结构的优化和权值学习

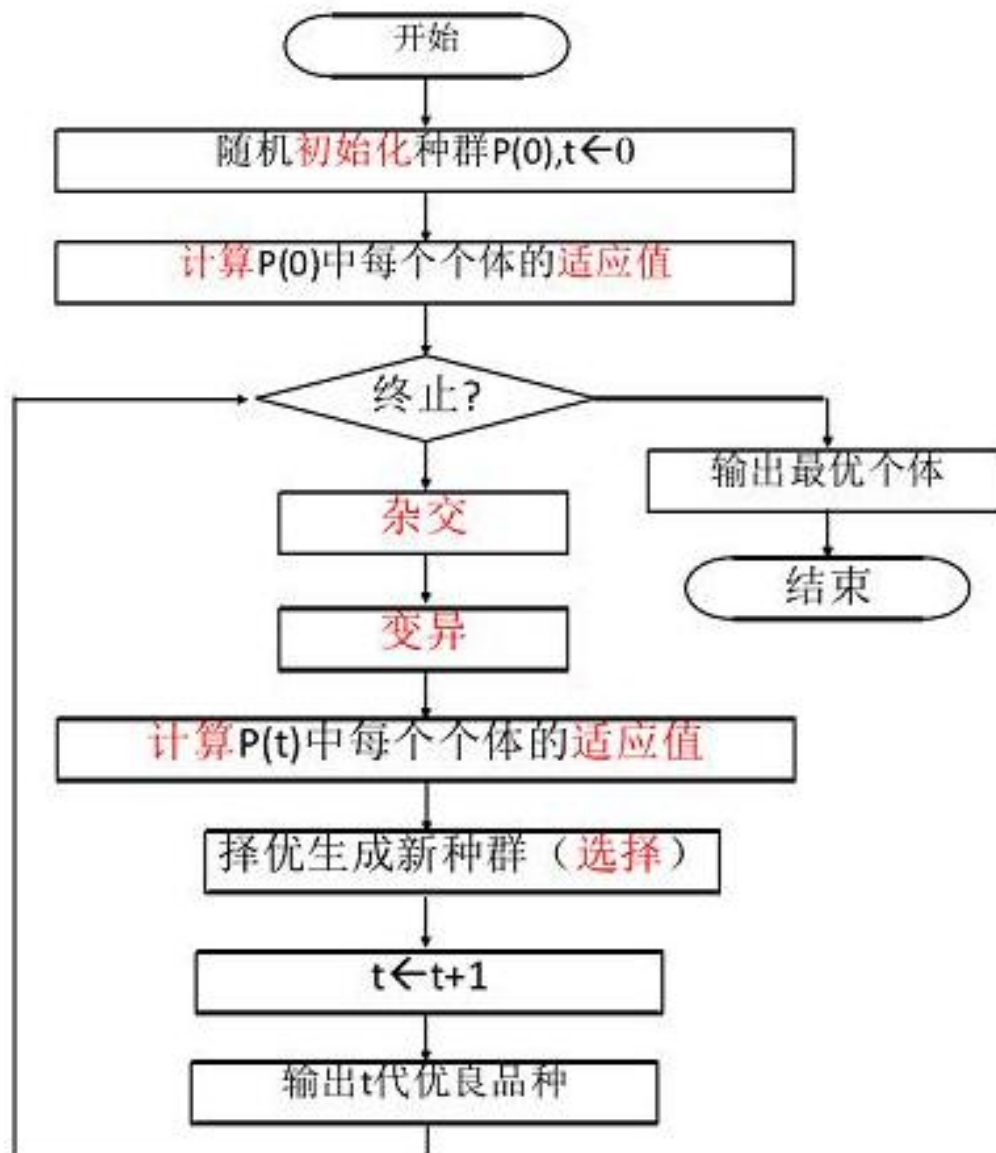
● 机器人

移动机器人路径规划、运动轨迹规划、结构优化和行为协调

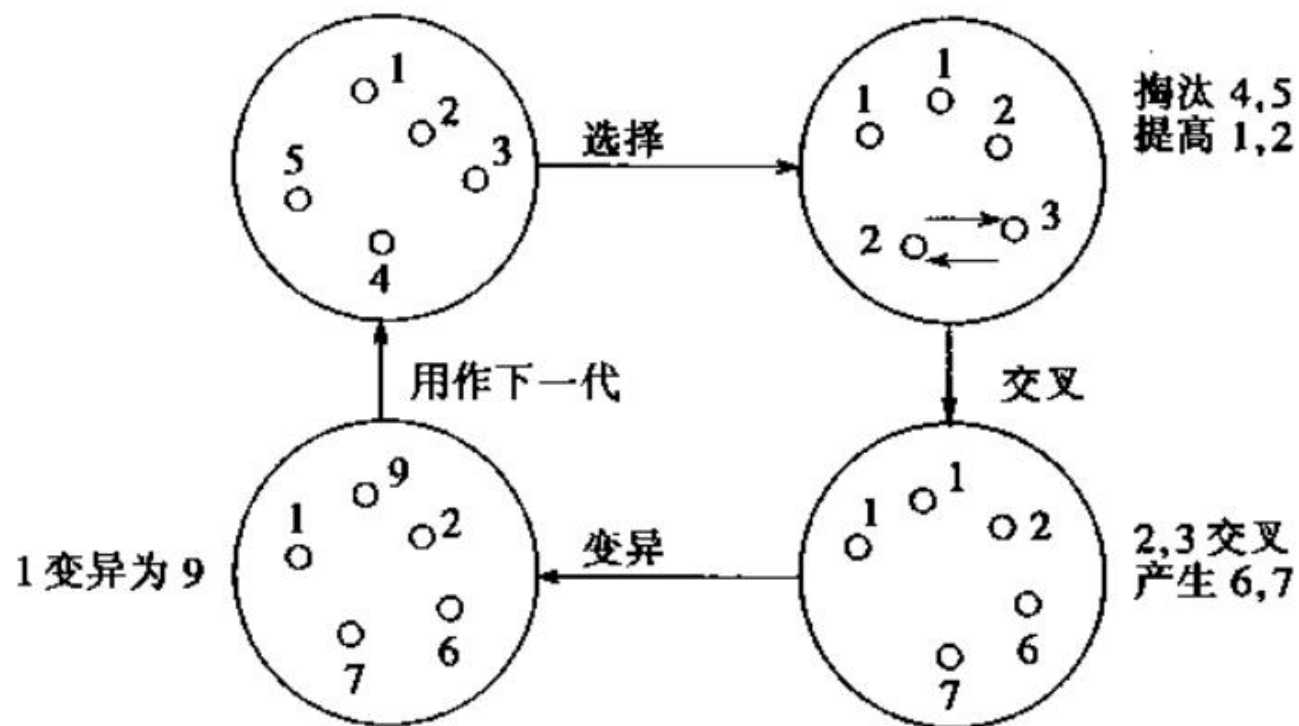
● 图像处理

图像处理过程中的特征提取、图像分割等的优化计算

3.1 遗传算法的基本思路



遗传算法执行过程



3.2 遗传算法基本概念

个体与种群(群体)

- 个体：模拟生物个体，个体也即可行解，对应染色体
- 种群(population)：模拟生物种群，由若干个体组成的群体，可行解集。

染色体与基因

染色体（**chromosome**）：

可行解的编码表示（**二进制编码**）。

可行解编码的分量，称为基因（**gene**）。

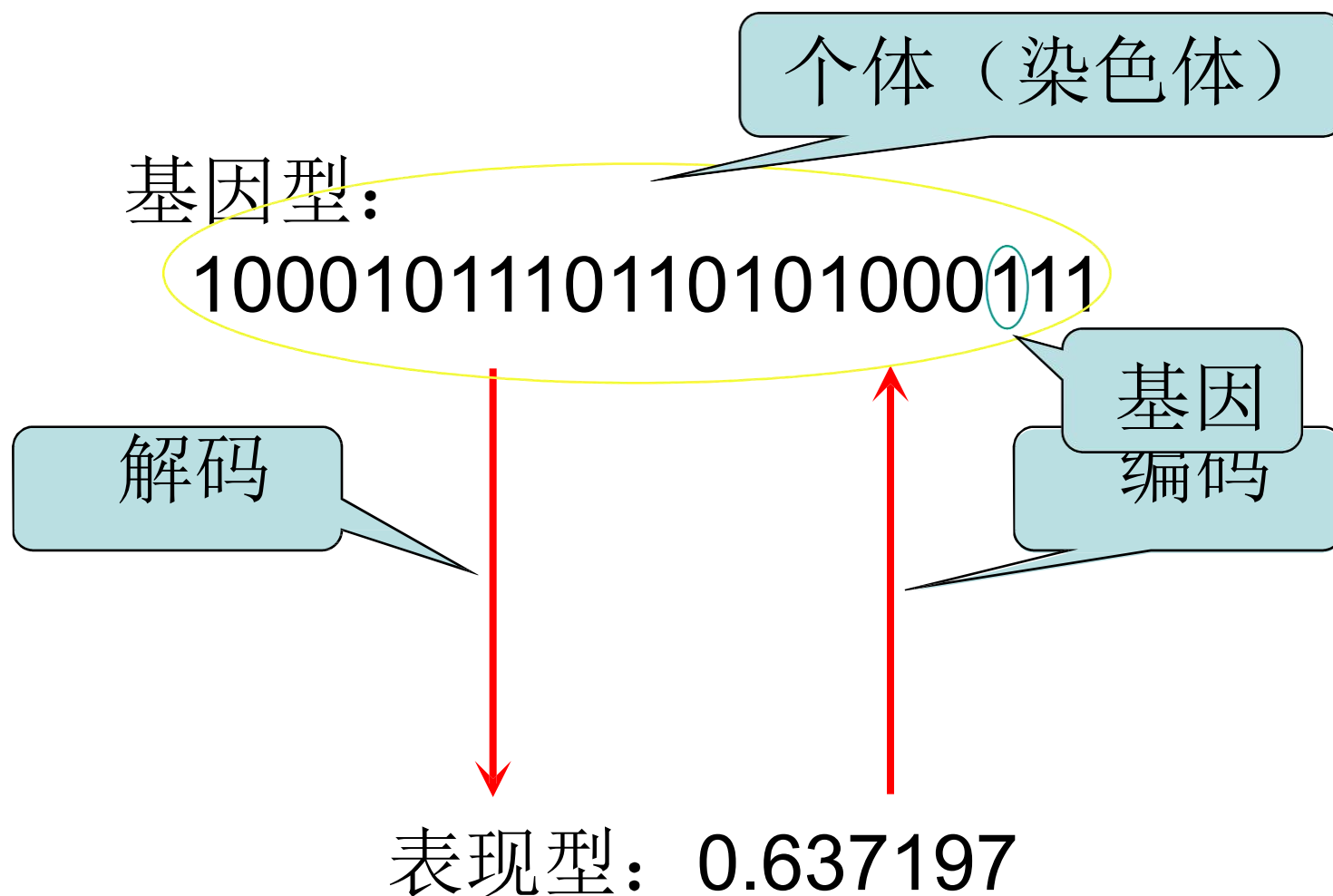
例如：

个体		染色体
9	-----	1001
(2, 5, 6)	-----	010 101 110

编码

通过某种机制把求解问题抽象为由特定符号排成的串。

二进制串进行编码是常见的方法



二进制编码：由0，1组成

将一个二进制串 $(b_n b_{n-1} \dots b_0)$ 转化为10进制数：

$$(b_n b_{n-1} \dots b_0)_2 = \left(\sum_{i=0}^n b_i \cdot 2^i \right)_{10} = x'$$

$$\begin{aligned} 1110 &= 0 \cdot 2^0 + 1 \cdot 2^1 + 1 \cdot 2^2 + 1 \cdot 2^3 \\ &= 0 + 2 + 4 + 8 \\ &= 14 \end{aligned}$$

举个例子把 $(13)_{10}$ 转化为二进制，我们需要做的就是

然后自下而上写出来 $(1101)_2$

$$\begin{aligned} \frac{13}{2} &= 6 \dots\dots 1 \\ \frac{6}{2} &= 3 \dots\dots 0 \\ \frac{3}{2} &= 1 \dots\dots 1 \\ \frac{1}{2} &= 0 \dots\dots 1 \end{aligned}$$

判断以下编码的合法性

(1) [1 0 2 0 1 1 0]--

(2) [1 0 1 1 0 0 1]---

(3) [0 1 1 0 0 1 0]--

(4) [0 0 0 0 0 0 0]-

(5) [2 1 3 4 5 7 6]--

(1) [1 0 2 0 1 1 0]-----不合法，不能出现“2”↵

(2) [1 0 1 1 0 0 1]-----合法↵

(3) [0 1 1 0 0 1 0]-----合法↵

(4) [0 0 0 0 0 0 0]-----合法↵

(5) [2 1 3 4 5 7 6]-----不合法，不能出现 0、1 以外的数字↵

适应度与适应度函数

- 适应度(fitness): 生物个体对环境的适应能力。
- 适应度函数(fitness function): 遗传算法用来评价个体(解)优劣的数学函数。

适应度函数值越大，解的质量越好

遗传操作

亦称遗传算子(genetic operator)，关于染色体的运算。

遗传算法中有三种遗传操作：

- 选择(selection)
- 交叉(crossover，亦称交换、交配或杂交)
- 变异(mutation，亦称突变)

选择算子

遗传算法使用**选择运算**实现对个体进行优胜劣汰操作：

- ① 适应度高的个体被遗传到下一代群体中的概率大；
- ② 不产生新个体

经典的选择算子采用**轮盘赌选择方法**。

轮盘赌选择方法

轮盘赌选择又称比例选择算子，

基本思想是：

个体被选中的概率与其适应度函数值大小成正比。

设群体大小为 n ，个体 i 的适应度为 F_i ，则个体 i 被选中遗传到下一代群体的概率为：

$$p_i = \frac{F_i}{\sum_{i=1}^n F_i}$$

轮盘赌选择方法的实现步骤

轮盘赌选择法可用如下过程模拟来实现：

- ① 在 $[0,1]$ 内产生一个均匀分布的随机数 r 。↓
- ② 若 $r \leq q_1$ ，则染色体 x_1 被选中。↓
- ③ 若 $q_{k-1} < r \leq q_k$ ($2 \leq k \leq N$)，则染色体 x_k 被选中。↓
- ④ 其中的 q_i 称为染色体 x_i ($i = 1, 2, \dots, N$) 的积累概率，其计算公式为：↵

$$q_i = \sum_{j=1}^i p(x_j) \quad \leftarrow$$

假若有染色体：↵

$$s_1 = 13 \text{ (01101)} \quad s_2 = 24 \text{ (11000)} \quad s_3 = 8 \text{ (01000)} \quad s_4 = 19 \text{ (10011)} \quad \leftarrow$$

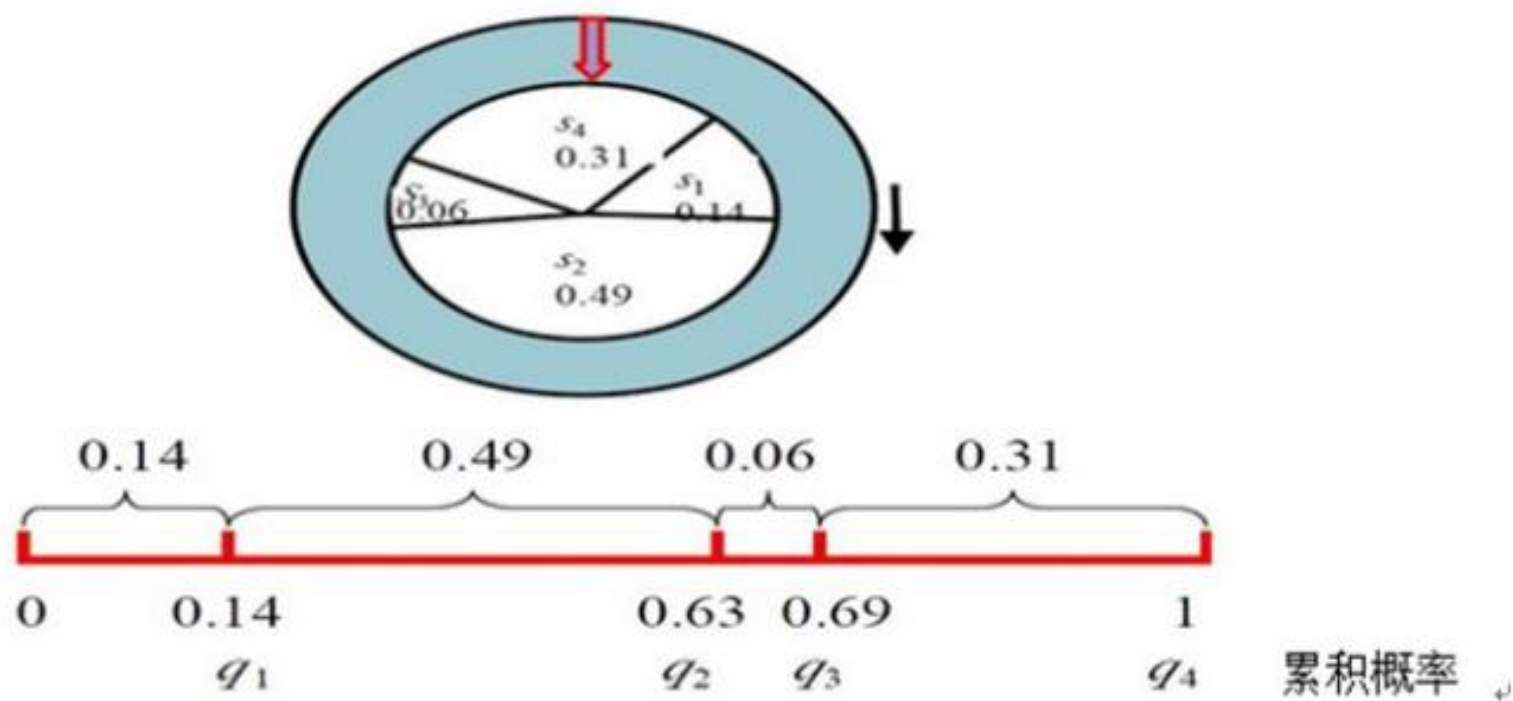
若要求解 $f(x) = x^2$ 的适应度，那么轮盘算法如下：↵

$$f(s_1) = x_1^2 = 169 \quad f(s_2) = 576 \quad f(s_3) = 64 \quad f(s_4) = 361 \quad \leftarrow$$

则染色体的选择概率为：↵

$$p(s_1) = 0.14 \quad p(s_2) = 0.49 \quad p(s_3) = 0.06 \quad p(s_4) = 0.31 \quad \leftarrow$$

即像抽奖中的轮盘一样：↵



例如设从区间 $[0,1]$ 中产生 4 个随机数：↵

$$r_1 = 0.450126, \quad r_2 = 0.110347, \quad r_3 = 0.572496, \quad r_4 = 0.98503 \quad \text{↵}$$

染色体的选择概率有：



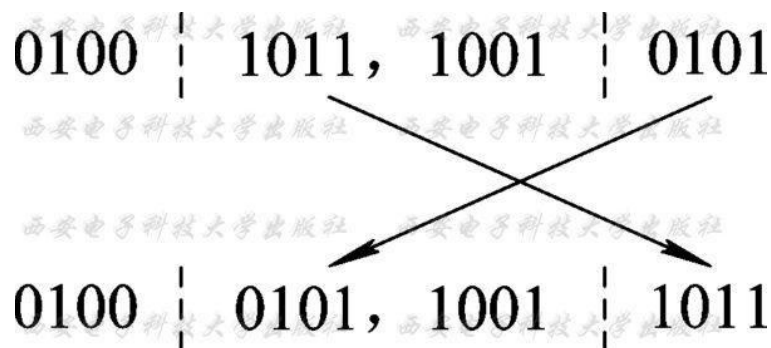
染色体	适应度	选择概率	累积概率	选择次数
13 (01011)	169	0.14	0.14	1
24 (11000)	576	0.49	0.63	2
8 (01000)	64	0.06	0.69	0
19 (10011)	361	0.31	1	1

可以看到概率越大，适应度越高，被选择的概率就越大。

交叉

互换两个染色体某些位上的基因, 是产生新个体的主要方法

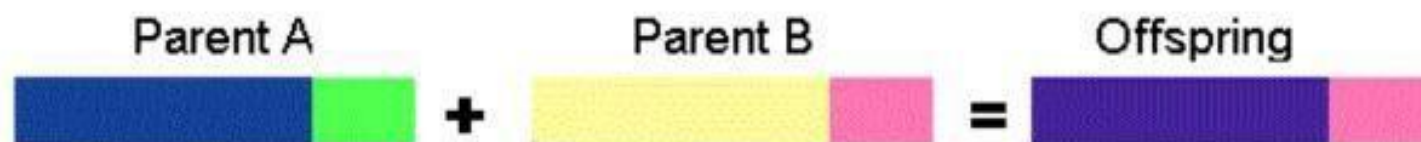
例如, 设染色体 $s_1=01001011$, $s_2=10010101$,
交换其后4位基因, 即



$$s_1'=01000101, \quad s_2'=10011011$$

可以看做是原染色体 s_1 和 s_2 的子代染色体。

Crossover可以发生在一个断点，如下图



$$11001011 + 11011111 = 11001111$$

One Point Crossover

两个断点交叉互换，其至多个断点交叉互换



$$11001011 + 11011111 = 11011111$$

Two Point Crossover

变异

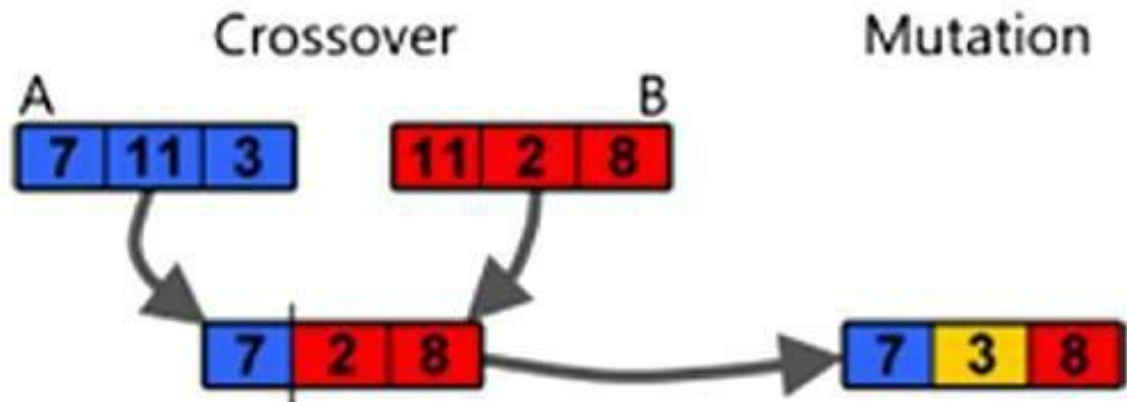
改变染色体某个(些)位上的基因。

例如, 设染色体 $s=11001101$, 将其第三位上的0变为1,

即

$$s=11\underline{0}01101 \rightarrow 11\underline{1}01101 = s'.$$

s' 也可以看做是原染色体 s 的子代染色体。

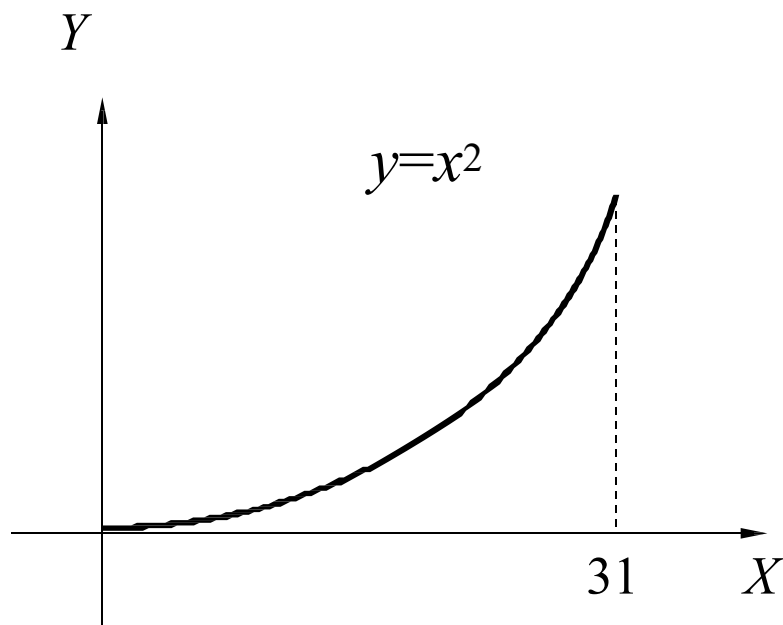


生物遗传概念在遗传算法中的对应关系

生物遗传概念	遗传算法中的作用
适者生存	在算法停止时，最优目标值的解有最大的可能被保留
个体(individual)	解
染色体(chromosome)	解的编码（字符串，向量等）
基因(gene)	解中每一分量的特征（如各分量的值）
适应性(fitness)	适应函数值
群体(population)	选定的一组解（其中解的个数为群体的规模）
种群(reproduction)	根据适应函数值选取的一组解
交配(crossover)	通过交配原则产生一组新解的过程
变异(mutation)	编码的某一个分量发生变化的过程

4 遗传算法应用举例

例1 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

- ① 原问题可转化为在区间 $[0, 31]$ 中搜索能使 y 取最大值的点 a 的问题。
- ② 则 $[0, 31]$ 中的点 x 即是个体,
- ③ 函数值 $f(x)$ 恰好可以作为 x 的适应度函数, 区间 $[0, 31]$ 就是一个(解)空间。
- ④ 因此, 给出个体 x 的适当染色体编码, 该问题即可用遗传算法来解决。

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位二进制数编码个体;

取下列个体组成初始种群 S_1 :

$$s_1 = 13 \text{ (01101)}, \quad s_2 = 24 \text{ (11000)}$$

$$s_3 = 8 \text{ (01000)}, \quad s_4 = 19 \text{ (10011)}$$

(2) 定义适应度函数, 取适应度函数: $f(x) = x^2$

(3) 计算各代种群中的各个体的适应度, 并对其进行遗传操作, 直到适应度最高的个体出现为止。

(即31 (11111))

首先计算种群 S_1 中各个体的适应度 $f(s_i)$

$$s_1 = 13(01101), \quad s_2 = 24(11000)$$

$$s_3 = 8(01000), \quad s_4 = 19(10011)$$

。 易求得

$$f(s_1) = f(13) = 13^2 = 169$$

$$f(s_2) = f(24) = 24^2 = 576$$

$$f(s_3) = f(8) = 8^2 = 64$$

$$f(s_4) = f(19) = 19^2 = 361$$

再计算种群 S_1 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

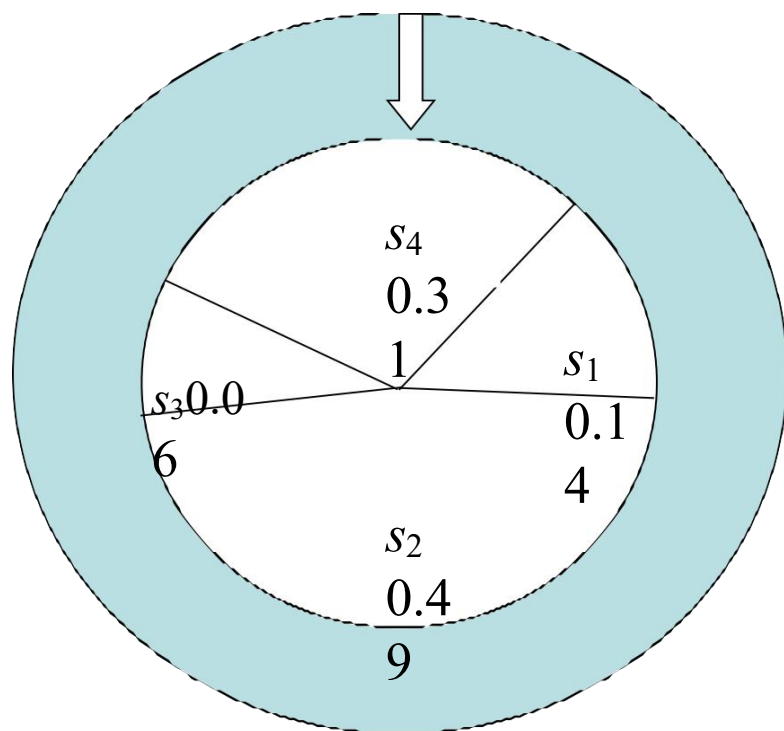
$$P(s_1) = P(13) = 0.14$$

$$P(s_2) = P(24) = 0.49$$

$$P(s_3) = P(8) = 0.06$$

$$P(s_4) = P(19) = 0.31$$

赌轮选择法



赌轮选择示意

- ① 在 $[0, 1]$ 区间内产生一个均匀分布的随机数 r 。
 - ② 若 $r \leq q_1$, 则染色体 x_1 被选中。
 - ③ 若 $q_{k-1} < r \leq q_k$ ($2 \leq k \leq N$), 则染色体 x_k 被选中。
- 其中 q_i 称为染色体 x_i ($i=1, 2, \dots, n$) 的积累概率, 其计算公式为

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

选择-复制

设从区间 $[0, 1]$ 中产生4个随机数如下:

$$r_1 = 0.450126, \quad r_3 = 0.572496,$$

$$r_2 = 0.110347, \quad r_4 = 0.98503$$

染色体	适应度	选择概率	积累概率	选中次数
$s_1=01101$	169	0.14	0.14	1
$s_2=11000$	576	0.49	0.63	2
$s_3=01000$	64	0.06	0.69	0
$s_4=10011$	361	0.31	1.00	1

于是, 经复制得群体 S_1 :

$$s_1' = 11000 \quad (24), \quad s_2' = 01101 \quad (13)$$

$$s_3' = 11000 \quad (24), \quad s_4' = 10011 \quad (19)$$

交叉

设交叉率 $p_c=100\%$ ，即 S_1 中的全体个体都参加交叉运算。

设 s_1' 与 s_2' 配对， s_3' 与 s_4' 配对。

分别交换后两位基因，得新染色体集 S_2 ：

$$s_1' = 11001 \text{ (25)}, \quad s_2' = 01100 \text{ (12)}$$

$$s_3' = 11011 \text{ (27)}, \quad s_4' = 10000 \text{ (16)}$$

群体 S_1 ：

$$s_1' = 11000 \text{ (24)}, \quad s_2' = 01101 \text{ (13)}$$

$$s_3' = 11000 \text{ (24)}, \quad s_4' = 10011 \text{ (19)}$$

变异

设变异率 $p_m=0.001$ ， 则群体 S_1

共有

$$5 \times 4 \times 0.001 = 0.02 \text{ 位基因}$$

可以变异。

0.02位显然不足1位，所以本轮遗传操作不做变异。

于是，得到第二代种群**S**：

$s_1=11001$ (25) , $s_2=01100$ (12)

$s_3=11011$ (27) , $s_4=10000$ (16)

第二代种群 S_2 中各染色体的情况

染色体	适应度	选择概率	积累概率	估计的选中次数
$s_1=11001$	625	0.36	0.36	1
$s_2=01100$	144	0.08	0.44	0
$s_3=11011$	729	0.41	0.85	2
$s_4=10000$	256	0.15	1.00	1

$$r_1 = 0.450126, \quad r_3 = 0.572496,$$

$$r_2 = 0.110347, \quad r_4 = 0.98503$$

假设这一轮选择-复制操作中，新种群 S 中的
4个染色体都被选中，则得到群体：

$$s_1' = 11001 \text{ (25)}, \quad s_2' = 01100 \text{ (12)}$$

$$s_3' = 11011 \text{ (27)}, \quad s_4' = 10000 \text{ (16)}$$

做交叉运算，让 s_1' 与 s_2' ， s_3' 与 s_4' 分别交换后三基因，得

$$s_1' = 11100 \text{ (28)}, \quad s_2' = 01001 \text{ (9)}$$

$$s_3' = 11000 \text{ (24)}, \quad s_4' = 10011 \text{ (19)}$$

这一轮仍然不会发生变异。

于是，得第三代种群S:

$$s_1=11100 \text{ (28)}, s_2=01001 \text{ (9)}$$

$$s_3=11000 \text{ (24)}, s_4=10011 \text{ (19)}$$

第三代种群S中各染色体的情况

染色体	适应度	选择概率	积累概率	估计的选中次数
$s_1=11100$	784	0.44	0.44	2
$s_2=01001$	81	0.04	0.48	0
$s_3=11000$	576	0.32	0.80	1
$s_4=10011$	361	0.20	1.00	1

$$r_1 = 0.450126, \quad r_3 = 0.572496,$$

$$r_2 = 0.110347, \quad r_4 = 0.98503$$

设这一轮的选择-复制结果为：

$$s_1' = 11100 \text{ (28)}, \quad s_2' = 11100 \text{ (28)}$$

$$s_3' = 11000 \text{ (24)}, \quad s_4' = 10011 \text{ (19)}$$

做交叉运算，让 s_1' 与 s_4' ， s_2' 与 s_3' 分别交换后两位基因，得

$$s_1'' = 11111 \text{ (31)}, \quad s_2'' = 11100 \text{ (28)}$$

$$s_3'' = 11000 \text{ (24)}, \quad s_4'' = 10000 \text{ (16)}$$

这一轮仍然不会发生变异。

于是，得第四代种群S:

$$s_1=11111 \text{ (31)}, \quad s_2=11100 \text{ (28)}$$

$$s_3=11000 \text{ (24)}, \quad s_4=10000 \text{ (16)}$$

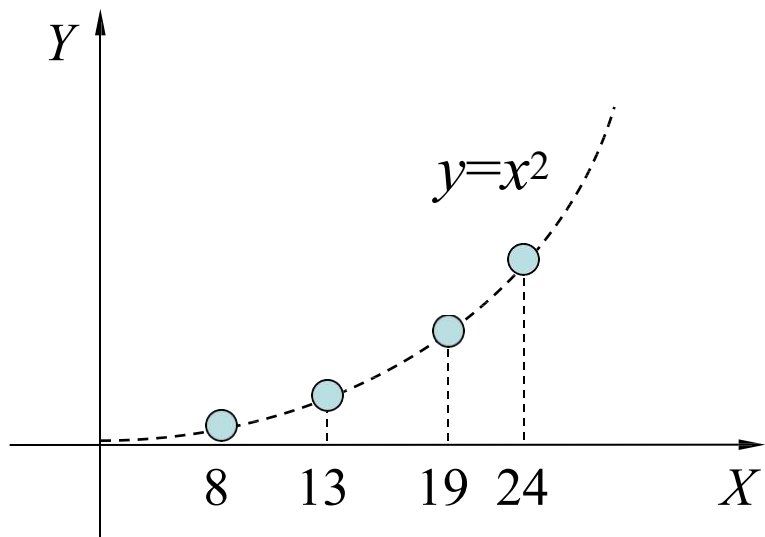
显然，在这一代种群中已经出现了适应度最高的个体 $s_1=11111$ 。

则遗传操作终止， 将个体“11111”作为最终结果输出。

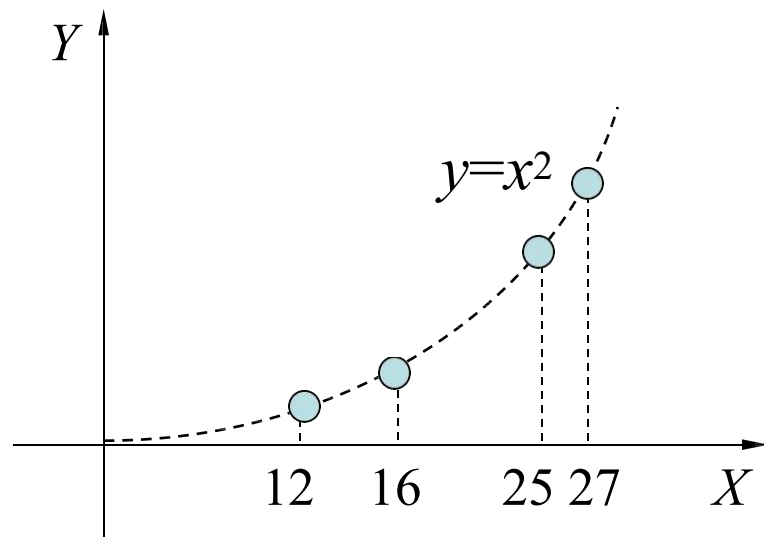
将染色体“11111”解码， 即得所求的最优解：31。

将31代入函数 $y=x^2$ 中， 即得原问题的解，

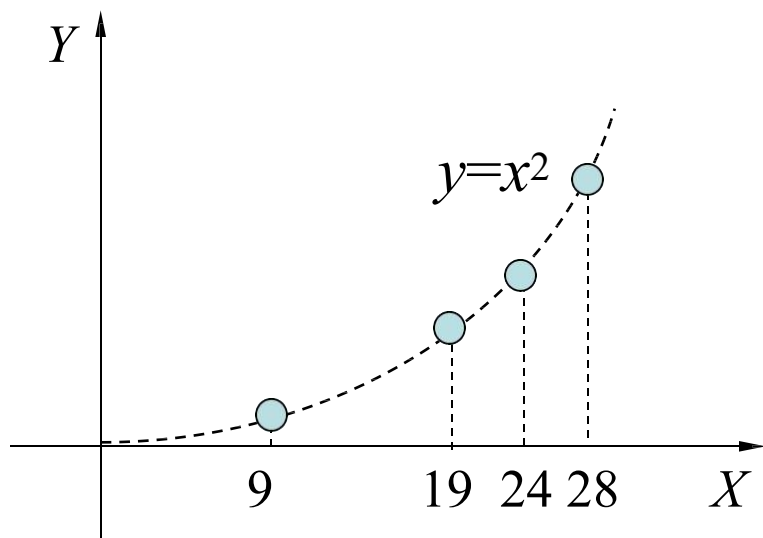
即 函数 $y=x^2$ 的最大值为961。



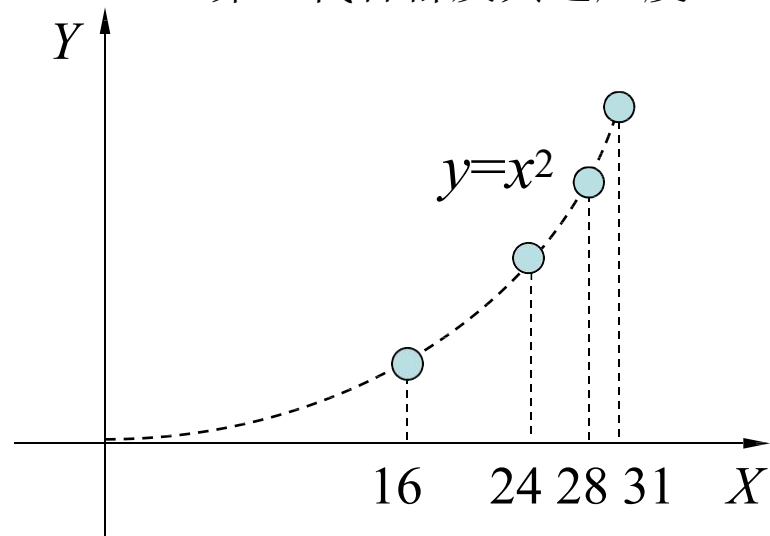
第一代种群及其适应度



第二代种群及其适应度



第三代种群及其适应度



第四代种群及其适应度

4 遗传算法主要步骤

- 首先初始化个体与种群
- 其次对优化问题的解进行编码
- 第三是适应函数的构造和应用
- 第四是染色体的结合:遗传算子（选择、交叉、变异）
- 最后运行参数。

运行参数

- (1) M : 种群规模
- (2) T : 遗传运算的终止进化代数
- (3) P_c : 交叉概率
- (4) P_m : 变异概率

基本遗传算法

- 步1 给定种群规模 N ，交叉率 P_c 和变异率 P_m ，适应度函数 $f(x)$ ，代数 T ；
- 步2 随机产生 N 个个体 s_1, s_2, \dots, s_N ，组成初始种群 $S=\{s_1, s_2, \dots, s_N\}$ ，置计数器 $t=1$ ；
- 步3 计算 S 中每个个体的适应度 $f()$ ；
- 步4 若终止条件满足，则取 S 中适应度最大的个体作为所求结果，算法结束。

-
- 步5** 按**选择概率** $P(x_i)$ ，从 S 中随机选定1个个体进行复制 N 次，组成群体 S_1
- 步6** 按**交叉率** P_c 所得个体数 c ，从 S_1 中随机确定 c 个染色体，配对进行交叉操作，得群体 S_2 ；
- 步7** 按**变异率** P_m 所决定的变异次数 m ，从 S_2 中随机确定 m 个染色体，进行变异操作，得群体 S_3 ；
- 步8** 将群体 S_3 作为新一代种群，即用 S_3 代替 S ， $t = t+1$ ，转步3

遗传算法基本流程图

