### 构成要素

遗传算法是对生物遗传和进化过程中选择、交叉、变异机理的模仿，来完成对问题最优解的自适应搜索过程。

遗传算法主要有四个构成要素，

第一个，染色体编码方式。基本遗传算法使用固定的二进制符号串来表示群体中的个体。我们最广泛使用的就是01二进制的编码方式。

第二是个体适应度评价，按照个体适应度来决定当前群体中每个个体遗传到下一代群体中机会的多少。

第三个，遗传算子，包括了选择、交叉和变异运算。分别模拟自然界对生物个体的筛选、个体之间繁衍和变异的现象。

最后是确定群体的规模、进化代数以及交叉变异可能性的大小。

接下来，我们将详细讲解这四个构成要素。

首先是染色体编码方式。在大自然中，优胜劣汰，我们将适应大自然的个体类推为可行解，这就是对个体编码的操作。编码方式决定了个体染色体的排列方式，也决定了个体的解码方式。

### 编码原则

一个好的编码方法能够极大地提高算法的实现和执行，产生最优解。为了提高遗传算法运行效率，提出了两条操作性较强的实用编码原则，第一是有意义的积木块编码原则，这个原则旨在使用易于生成适应度高的个体。

第二是最小字符集编码方式，这个原则说明了二进制编码的普适性，与其他编码方式相比，二进制编码方式能包含更多的处理模式。

### 四种编码方式

这两种编码方式只是一个指导大纲，并不适用所有问题，如今，人们已经提出了许多不同的编码方案，大致可以分为四大类：二进制、格雷码、浮点数和字符。

二进制编码方式是遗传算法中最常见的一种编码方式，使用二进制符号01组成。二进制编码方式有很多优点，首先，二进制编码和解码操作简单，其次，交叉变异的遗传操作很容易实现。但它也有一个显著的确定，二进制编码符号串的长度与问题的求解精度有关，因此在遇到高精度问题时，我们会采用其他编码方式，这个我们后面会讲到。

格雷码编码方式指的是让连续两个整数所对应的编码值之间仅仅只有一个码位不同，这样一个差值成为一个海明距离。格雷码编码方式其实是二进制编码方式的一种变形，它能提高遗传算法是局部搜索能力，同时保留着一些二进制编码方式的优点。

当我们遇到多维、高精度要求的最优化问题求解时，采用二进制编码会大大增加编码串长度，比如为了表示小数点后5位数的数值，我们需要将数值放大10万倍，再用二进制表示，这样使得搜索空间变大，运行性能变差。这个时候有人提出了浮点数编码方式，也叫做真值编码方式。

最后是符号编码方式，当我们需要表达一个无数值意义，而有代码意义的符号集时，可以使用符号编码方式，比如我们常见的旅行商问题，旅行路线是可行解，城市没有数值意义，可以用符号表示。

### 适应度函数

在自然界中，对生存环境适应度高的个体将有更多的繁殖机会，而对环境适应度低的，繁殖机会就相对较少。与此类似，我们需要一个适应度的概念，找出个体中适应度较高的遗传下去，而适应度较低的个体遗传下去的机率相对较小。度量个体适应度的函数成为适应度函数。

适应度函数主要有两种求解，全局最大值和全局最小值。他们所对应的目标函数到搜索空间对应个体的适应度函数值转换关系如下。

### 适应度尺度变换

在遗传算法中，有些算法收敛得快，有些算法收敛得慢。比如在算法初期阶段，群体中少数几个个体的适应度相对其他个体来说非常高，这样在下一代群体中这些少数个体将占据很高的比例，群体的多样性降低，导致早期收敛，停留在局部最优解。为了克服这种现象，在算法运行初期阶段，要对一些适应度高的个体进行控制，从而限制其复制数量，维护群体的多样性。

反之，在算法初期，个体之间的适应度差异不大，无竞争力，导致算法运行效率降低，我们需要对个体的适应度进行适当放大，扩大最佳个体适应度与其他个体之间的差异，提高个体之间的竞争性。

目前我们有三种主要的适应度尺度变换方法。

### 选择算子

模拟自然界的优胜劣汰，我们使用选择算子，或称复制算子对群体中的个体进行操作。选择操作就是用来确定如何从父代群体中按某种方式选择哪些个体遗传到下一代群体中的一种遗传运算。

### 常见选择算子

最常见的选择算子是基本遗传算法中的比例选择算子 。其基本思想是：各个个体被选中的概率与其适应度大小成正比，每个个体的适应度除以全部个体适应度总和，则为个体选中的概率。

最优保存策略则是将当前群体中适应度最高的个体直接保留到下一代，不进行交叉运算和变异运算。并替换掉群体中适应度最低的个体。但这个算法的全局搜索能力不强，容易陷入局部最优解。

确定式采样选择的操作过程是：先计算出期望生存数目，将个体进行降序排序，选取期望生存个数的个体，加入下一代群体中。

除此之外，还有无回放随机选择、无回放余数随机选择、排序选择和随机联赛选择等方式。

### 交叉算子

遗传算法中的交叉运算，指对连个互相配对的染色体按某种方式交换其部分基因，从而形成两个新的个体。交叉算子的设计主要包括两个方面：1确定交叉的位置，2进行部分基因交换。

### 常见交叉算子

最常见的交叉算子是单点交叉，也成为简单交叉。随机设置一个交叉点，互相交换个体的部分染色体。

双点交叉和多点交叉 则是随机设置两个或多个交叉点。操作方法与单点交叉类似。

均匀交叉则是先随机产生一个与编码串长度等长的屏蔽字，如果屏蔽字为1，则交换对应的基因值，如果屏蔽字为0，则保留双方各自的基因值。均匀交叉也可以看作是多点交叉。

算术交叉是对两个个体的基因值进行线性组合运算。

### 变异算子

在自然界上，生物基因的转录过程中，会出现突变的情况。遗传算法也引入了变异算子来模拟这个环节，从而产生新的个体。变异算子能够改善遗传算法是局部搜索能力，维持群体多样性，防止出现早熟现象。

### 常见变异算子

第一种是基本位变异，指的是以某个概率随机在某一个或几个基因座上做变异运算。由于变异概率低，改变的只是一个或几个基因值，其发挥作用较慢，效果不明显。

因此，第二种变异算子：均匀变异则是依次指定个体编码串中的每个基因座为变异点，而不像基本位变异一样，变异的基因座是固定的。

为了使得对某个重点区域进行局部搜索，又提出了非均匀变异算子，非均匀变异算子随着迭代次数的增加，搜索区域越来越集中到最有希望的重点区域中。

高斯变异则是与非均匀变异算子类似，也是重点搜索区域的局部搜索方法。根据正态分布特性取搜索局部区域。

### 遗传算法步骤

以上讲完了遗传算法的基本原理，现在我们来总结遗传算法的步骤

1. 确定决策变量以及各种约束条件
2. 建立优化模型
3. 确定可行解的染色体编码方式
4. 确实解码方式
5. 确定个体适应度的量化评价方法
6. 设计遗传算子
7. 确定遗传算法的有关运行参数

最后是算法流程图，便于大家的理解。