从零开始学数学建模

遗传算法

主讲人：北海

b站/公众号：数学建模BOOM

遗传算法

- 算法简介
- 适用赛题
- 典型例题与原理讲解
- 代码求解

遗传算法

从做菜说起

优化类问题

- 北海是一名大厨，想要创造一道美味的菜肴
- 首先**随机**生成多个原始配方
- 每种配方所用的原料（鸭脖、鸡肉、大肠等）与手法（煎炒焖炸卤炖）组合不同
- 现实中考虑调料用量、烹饪时间等等变量，会有无穷多种解，传统算法难以求解



遗传算法

❑从做菜说起

❑ 适应度函数、交叉

- 请评委对几种配方做出的菜**打分**
- ★ **分数高的配方**进行**配方交叉**，保留一部分评分高的配方要素、舍弃评分低的配方
- 例如配方A和配方C的分数都高，A是卤鸭脖，C是炖大肠
- 配方交叉尝试新一组方案：“炖鸭脖”和“卤大肠”

遗传算法

❑从做菜说起

❑ 变异

- 有时会在配方交叉之后，再变更食材或烹饪方式
- 就像是在配方中**随机**使用了一些**与原配方无关**的调料或者做法（鸭脖改成鼠片）
- 变异可能带来惊喜（评分高），也可能有惊无喜（试试就逝世），所以只小概率进行

遗传算法

❑从做菜说起

❑ 迭代

- 再对新配方的菜评分，继续交叉、小概率变异……不断循环直至无改进空间为止

❑ 遗传算法思想

烹饪配方	生物学	遗传算法
多种配方	群体	可行解集合
单个配方	个体	可行解
配方内容(原料、方式等)	基因	可行解的分量
评分	适应度	适应度函数值
配方交叉	交叉	交叉操作
随机新操作（故意的）	变异	变异操作
做出美味佳肴	物种进化	最优解（近似）

❑ 本期课程第三部分“典型例题与原理讲解”会详细讲解遗传算法的原理与解题过程；第四部分讲解代码实现

遗传算法

- 算法简介
- 适用赛题
- 典型例题与原理讲解
- 代码求解

遗传算法

❑适用赛题

❑ 函数优化、组合优化问题

- 对于一些非线性、多模型、多目标的函数优化问题
- 不依赖于问题的背景领域，使用方便，连续/离散、单峰/多峰等等各种形式均可

❑ NP-Hard问题

- 模拟退火算法中讲过的TSP问题；背包问题、图形划分问题等
- 在NP-Hard问题方面普遍来说各类启发式算法均可

❑ 优缺点

- 相比模拟退火，相比良好的全局搜索能力，不易陷入局部最优
- 相比粒子群算法，常规的遗传算法可直接求解离散问题
- 缺点：由于变异是随机的，局部搜索能力差；相对其他算法更耗时、思路复杂抽象；

- 可多种算法结合改进，例如遗传算法优化神经网络等混合算法（新手慎用）

遗传算法

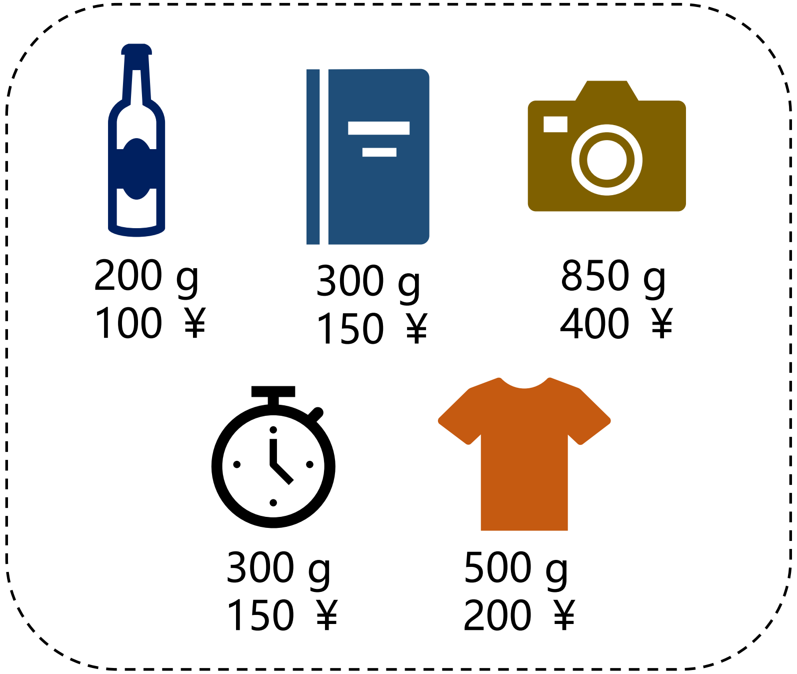
- 算法简介
- 适用赛题
- 典型例题与原理讲解
- 代码求解

遗传算法

❑0-1背包问题

❑ 组合优化

- 将一堆物品尽量塞进一个背包里，使得包内物品**总价值最高**
- 假设背包容量足够大，但有**载重上限**，物品太重会被撑破
- 物品**无法切割**，每个物品只能“放”与“不放”

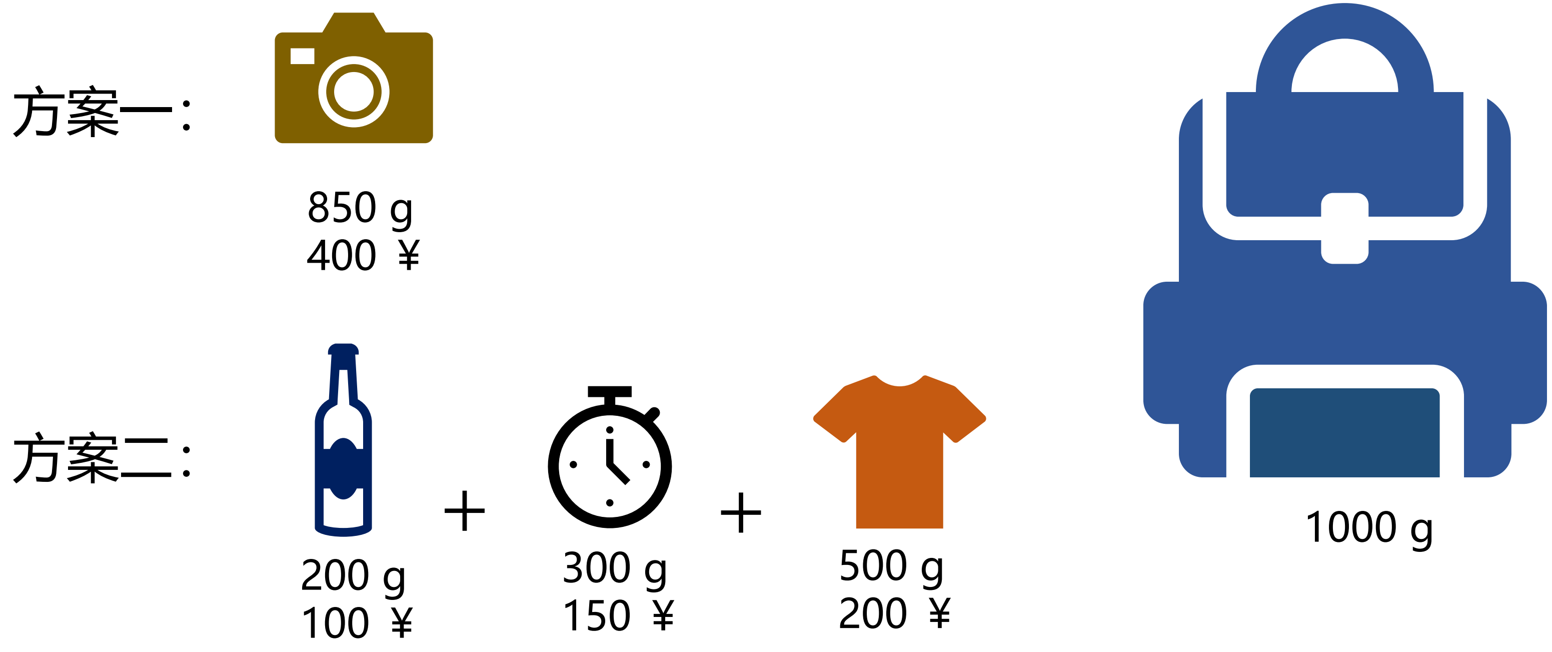


遗传算法

❑0-1背包问题

❑ 优先装价值大的?

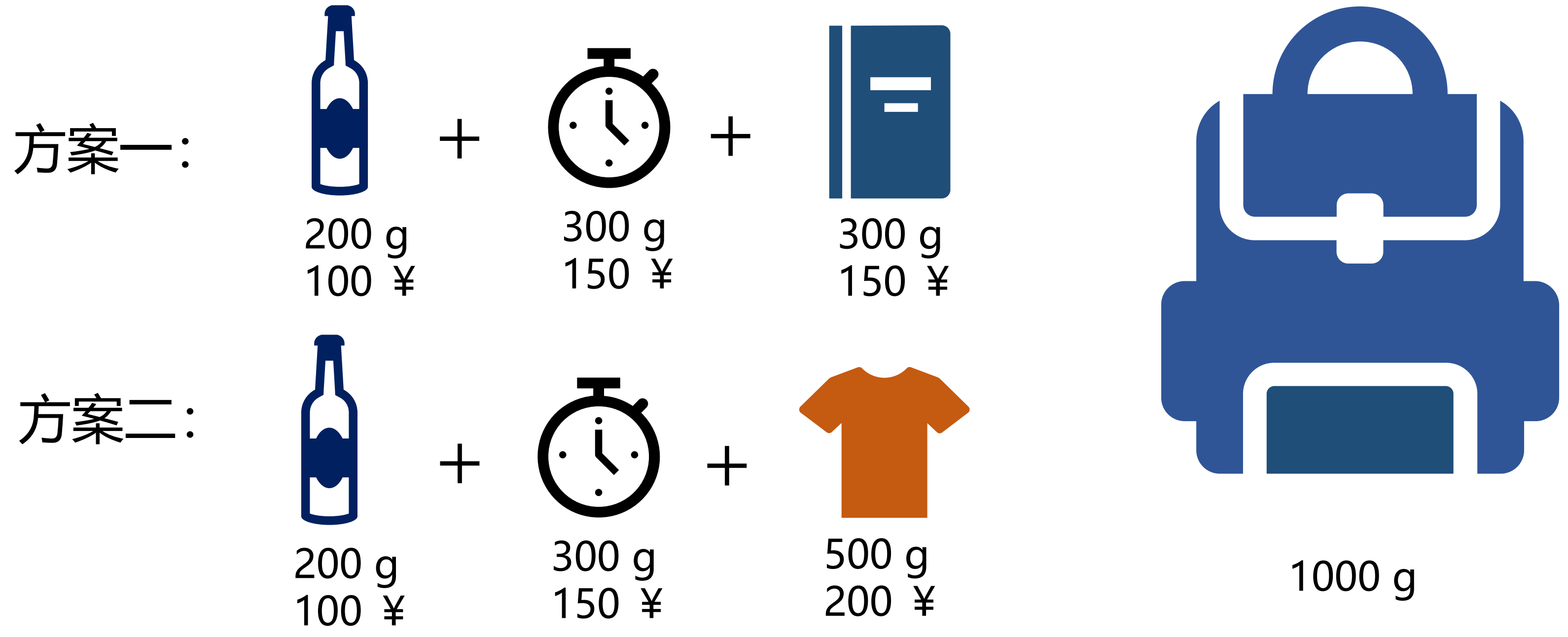
- 既然目标是“包内总价值最大”，那优先装价值高的物品?
- 但实际价值高的物品可能重量也很大，背包剩余容重量不足以再放其他物品
- 反而不如多塞几个的低重量低价值物品



❑0-1背包问题

❑ 优先装“价值/重量”高的?

- 基于上一种方法失效，有人会进一步想到，先装“价值/重量”的来最大化利用承重
- 五个物品的价值/重量：酒(0.5)=表(0.5)=书(0.5)>相机(0.47)>衣服(0.4)
- 该方案仍无法确保求得最优解
- 注意：如果物品可分割（随意切比例来装），那么该方法可求得最优解



遗传算法

❑0-1背包问题

❑ 穷举法罗列出所有可能的结果，找出最优?

- 假设先不考虑乘重上限，把5个物品装进背包，总共有多少可能的组合方案?
- 一个都不装：$C_5^0 = 1$种方案；装一个物品有$C_5^1 = 5$种方案；
- 装两个物品有C_5^2种方案；装三个物品有C_5^3种方案；
- 装四个物品有C_5^4种方案；装五个物品有C_5^5种方案；
- 总共$C_5^0 + C_5^1 + C_5^2 + C_5^3 + C_5^4 + C_5^5 = 2^5 = 32$种方案

- 如果总共有n个物品，那将是2^n种方案
- 传统方法无解或者运算量过大时，就该考虑启发式算法了

❑遗传算法求解0-1背包问题

❑ 例题：现有12份快递需要配送，背包容量350，每份快递的体积不同、收益不同，应该将哪些物品放进背包，使得所总体积不超过背包容量且总收益最大?

序号	1	2	3	4	5	6	7	8	9	10	11	12
体积	78	39	90	82	61	52	12	37	49	55	64	8
价值	64	37	22	41	76	34	22	21	10	57	32	12

- 用“0”和“1”表示物品的装包状态
- 一个物品装包状态为0代表没有被装进包中，1代表被装进包中
- 例如“100111001001”代表第1，4，5，6，9，12个物品被装进背包

- 每一个物品的装包状态（0或1）作为一个**基因**
- 问题的一个解：一组12个数构成的数组为一个**个体（染色体）**

遗传算法

❑编码与解码

❑ 问题与数学语言的转换

个体（染色体）	1	0	0	1	1	1	0	0	1	0	0	1
选择装包（问题的解）	物品1	物品2	物品3	物品4	物品5	物品6	物品7	物品8	物品9	物品10	物品11	物品12

容量有限的背包

遗传算法

❑算法思路

❑ 步骤一：种群初始化

- 有N个物品，则随机生成N个数（0或1）构成一个数组，作为一个个体，及可能的解
- 重复上一步，生成多个个体，构成种群（解集）
- 群体规模太小容易陷入局部最优，太大又会使计算复杂度高费时间，一般设置20到200

个体1	1	0	0	1	1	1	0	0	1	0	0	1
个体2	0	0	1	1	0	1	0	1	1	0	0	0
⋮												
个体N	1	1	1	0	1	0	0	1	0	0	0	0

基因：解分量

个体：解

种群：解集

遗传算法

❑算法思路

❑ 步骤二：选择运算

- 求每个个体的适应度：把解x带入目标函数求函数值$f(x)$
- 目的是从第t代群体$P(t)$中挑选出优良个体，把基因遗传到下一代群体$P(t+1)$

❑ 选择操作的准则

- 1、适应度越高的个体被选中的概率越大
- 2、适应度较低的个体仍有被选中的可能

❑ 为什么不直接从大到小排序，直接选适应度最高的几个个体进行后续操作?

- 适应度最高的几个个体，也意味着适应度高度相似
- 往往其基因（解分量）也很相似
- 如果每次都只选适应度最高的个体，意味着每轮迭代所选的个体相似度很高
- 那么后代的相似度也会很高，进化陷入停滞，意味着陷入局部最优解

遗传算法

❑算法思路

❑ 轮盘赌法

- 1、计算每个个体被选中概率$P(x_i)$，概率值与其适应度值成正比
- 2、计算每个个体对应的累积概率q_i，为从第1个个体到当前个体的选中概率之和

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

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

- 3、随机生成一个数组bet，其中的元素取值在0到1之间，并将元素从小到大排序
- ★ 4、若第i个个体的累积概率q_i大于$bet(i)$，则第i个个体被选中,并更新$bet(i)$为$bet(i+1)$否则选择第$(i+1)$个个体与$bet(i)$比较，直至选出一个个体为止
- 5、重复步骤4，直至选出与种群数量相等的个体数（其中有的个体被多次选中）

- （以上步骤较抽象，主要在代码讲解中看具体步骤来理解）

遗传算法

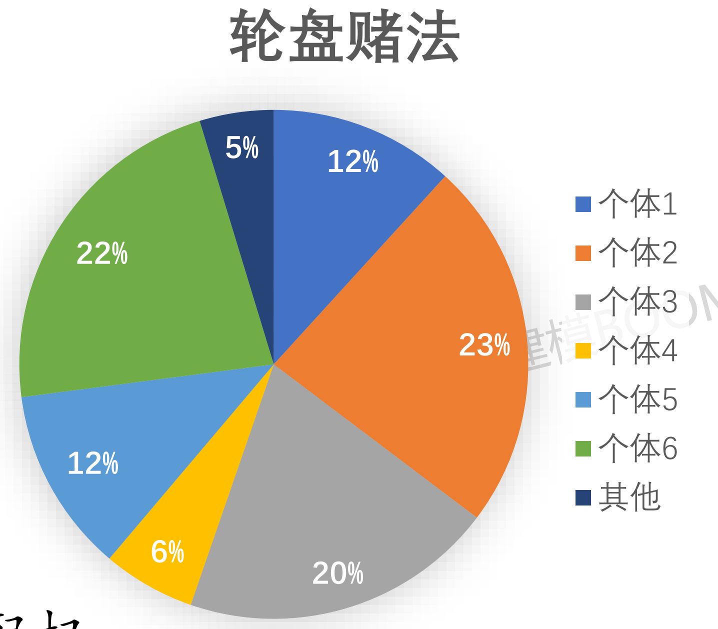
❑算法思路

❑ 轮盘赌法分析

- 第i个个体适应度值越大→被选中概率$P(x_i)$数值越大
- $P(x_i)$越大→ $P(x_i) = q_i - q_{i-1}$越大，即第i个个体对累积概率带来的“增大幅度”也越大
- ★ 增大幅度$(q_i - q_{i-1})$越大→新个体满足$q_i > bet$的概率也越大（回忆下bet是怎么变化的）
- 个体满足$q_i > bet(i)$即意味着被选中

- 结论：个体适应度值越大，被选中的概率越大

- 基因好、适应度大使得其对累积概率带来的“增幅”更大
- 类似在轮盘上该个体所占的面积越大，被选中的概率也越大
- 被选择的个体中会有重复
- 因为适应度高的个体被选中概率大而可能被选中多次
- 对应于生物界中基因优良生存能力强的个体可能具有多次交配权



遗传算法

❑算法思路

❑ 步骤三：交叉运算

- 在被选择的所有个体中，两个个体之间进行交叉操作
- 先根据交叉概率判断是否执行交叉操作，一般设置为80%到95%（即较大概率）
- 随机选择进行交叉的位置，例如随机选中第1、3、4、7、8位基因进行交叉：

交叉前的个体1	1	0	0	1	1	1	0	0	1	0	0	1
交叉前的个体2	1	1	1	0	1	0	0	1	0	0	0	0
交叉后的个体1	1	0	1	0	1	1	0	1	1	0	0	1
交叉后的个体2	1	1	0	1	1	0	0	0	0	0	0	0

遗传算法

❑算法思路

❑ 步骤四：变异运算

- 变异操作是为了利用“不确定性”来赌一把，或许更好，或许更差
- 先判断每个个体的每个基因是否进行变异运算，一般变异概率设为0.5%到5%即可
- 变异运算就是对变异的基因取反，0变1，1变0

变异前的个体

1	0	0	1	1	1	0	0	1	0	0	1

↓

变异后的个体

1	0	0	1	0	1	0	0	1	1	0	1

❑ 步骤五：迭代循环

- 选择交叉变异后得到新一代群体（解集），记录本轮迭代的最优个体和最优适应度
- 重复步骤二、三、四，直至满足终止条件（比如迭代满100次）

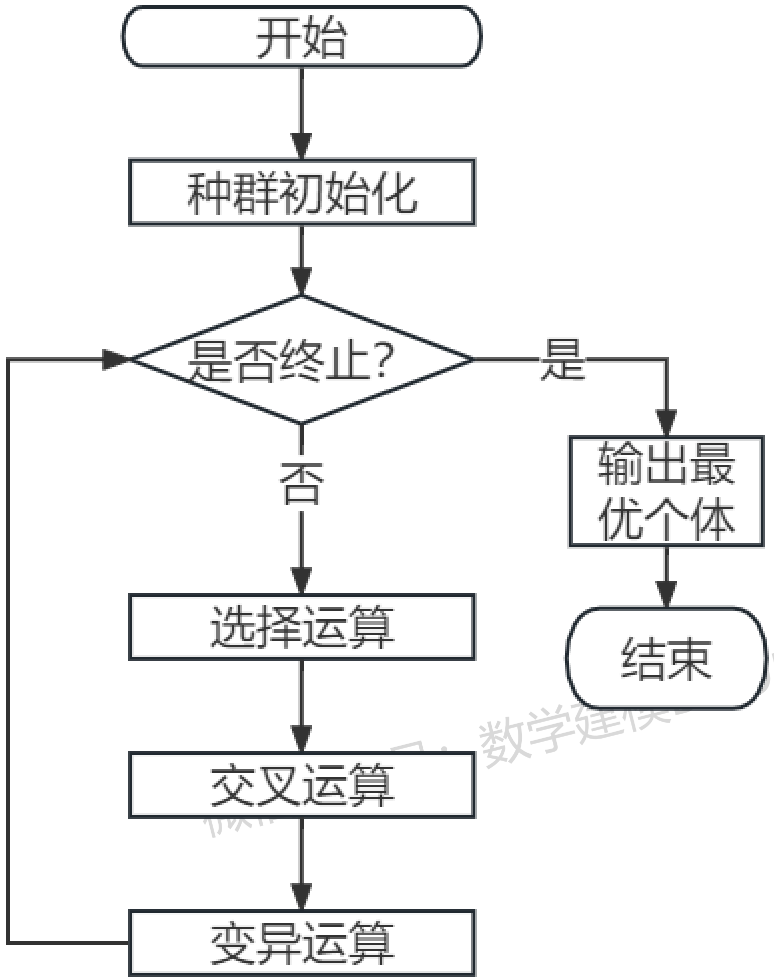
- *交叉和变异操作的具体实现方法不限于本课程所讲的，但基本思路和目的是一致的

遗传算法

❑算法思路

❑ 总结

背包问题	生物学	遗传算法
多种装包方案	群体	可行解集合
单个方案	个体	可行解
每个物体是否被选中	基因	可行解的分量
包内总价值	适应度	适应度函数值
两方案中的物品交换	交叉	交叉操作
随机改变物品的选择	变异	变异操作
包内总价值最大	物种进化	最优解（近似）



遗传算法

- 算法简介
- 适用赛题
- 典型例题与原理讲解
- 代码求解

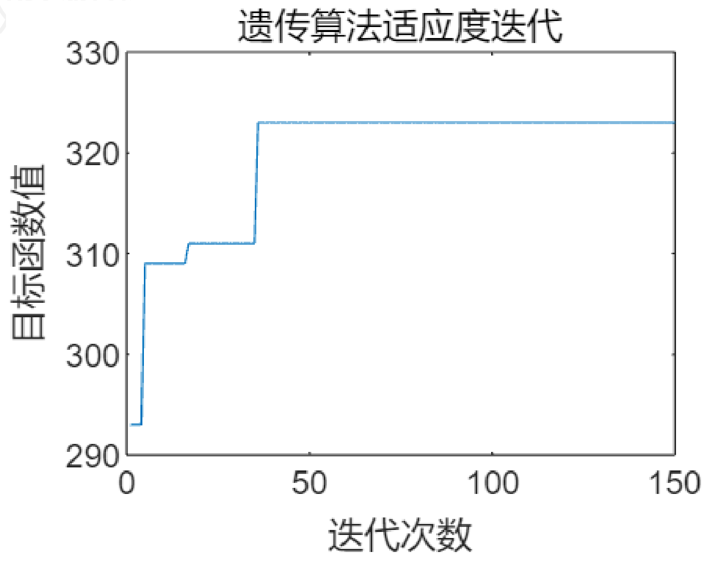
遗传算法

❑代码求解

❑ 接下来到MATLAB文件GA.mlx中讲解代码

- 求得的近似最优解，及最终的fBest：

1	1	0	0	1	1	1	1	0	1	0	1



- 即选择第1、2、5、6、7、8、10、12个商品装进背包
- 相应的近似最优函数值，及最终的maxfit：323
- 代码中多处涉及到随机概率运算，每次结果很可能不同
- 但种群的个体数量足够大、迭代次数足够多，近似最优函数值应该是相近的
- 最优解可能不止一个，因为可能多个解都对应这相同的函数值

遗传算法

❑写出你的笔记

❑ 费曼学习法

- 费曼学习法：以教代学
- 只有当你能够教会别人，才代表你真正学会了！

❑ 有奖征集：每学完一期课程，整理笔记，发布在各平台

- 将你每节课所学到的，整理出一套笔记
- 尽量不要照搬或截图课程的内容（必要的可截图）
- 可自行发布在知乎/CSDN等等各类平台

- 符合以下要求的文章，且文章点赞超过100或浏览量超1万的，可获取半价退款奖励（联系北海的QQ：1980654305 ）
- 1、标题设为：XXXX（模型或算法）——北海数学建模课程笔记
- 2、文章首行写：本文为北海的数模课程学习笔记，课程出自微信公众号：数学建模BOOM。

遗传算法

- “从零开始学数学建模”系列课程
  - 本期课程视频出自b站up：数学建模BOOM
  - 全套课程请关注微信公众号：数学建模BOOM，回复“课程”

END

