

# 统计推断在数模转换系统中的应用

组号 44

姓名：惠禹杰 学号 5130309124

姓名：徐开元 学号 5130309113

摘要：本文在阐述了遗传算法理论上，采用遗传算法对实验数据进行处理，通过数据分析证明了遗传算法的有效性，并对遗传算法进行了可行性的探讨。

关键词：统计推断，曲线拟合，遗传算法，插值，定标

## 1 引言

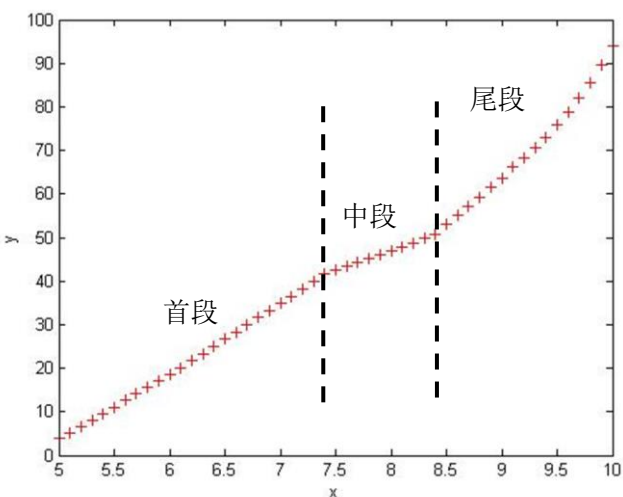
统计推断是一类重要的统计数学方法，主要包括参数估计与假设检验两个重要方面。由所给样本推断整体，是从数据中发现因果联系的重要步骤。客观规律的概率性质，也使统计推断成为科学概括的有力工具，使其在各种自然科学和社会科学研究、工程技术等领域都有较广泛的应用。而在统计推断方法便是统计推断的灵魂，对统计算法的灵活掌握与使用，将在工程实践或科学实验中取得事半功半的效果，所以在本次课程中，我们通过复杂而综合的数学建模、理论推算、算法制定、实验验证过程，对数模转换系统进行数据优化设计，使我们对统计推断算法有一定的了解，并验证了统计推断算法的有效性。

本体是由给定的某电子产品的输入输出值，来确定该电子产品的特性曲线。已知的是该曲线呈非线性，并有一下主要特征：

- Y 取值随 X 取值的增大而单调递增；
- X 取值在[5.0,10.0]区间内，Y 取值在[0,100]区间内；
- 不同个体的特性曲线形态相似但两两相异；
- 特性曲线按斜率变化大致可以分为首段、中段、尾段三部分，中段的平均斜率小于首段和尾段；
- 首段、中段、尾段单独都不是完全线性的，且不同个体的弯曲形态有随机性差异；

不同个体的中段起点位置、终点位置有随机性差异。

而下图也给出了大致的特性图线：



## 2 统计算法的理论基础

根据本课程的老师建议，我们在此使用遗传算法来进行问题的分析和解决。

遗传算法（Genetic Algorithm）是一类借鉴生物界的进化规律（适者生存，优胜劣汰遗传机制）演化而来的随机化搜索方法。其主要特点是直接对结构对象进行操作，不存在求导和函数连续性的限定；具有内在的隐并行性和更好的全局寻优能力；采用概率化的寻优方法，能自动获取和指导优化的搜索空间，自适应地调整搜索方向，不需要确定

的规则。遗传算法的这些性质，已被人们广泛地应用于组合优化、机器学习、信号处理、自适应控制和人工生命等领域。

在本问题中，可以笼统的概括为：编码，估值，选择，交叉，变异。

### 3 拟合方法的基本原理

#### 3.1 成本计算函数

根据所给的 469 组数据有效的曲线数据，我们认为大多数数据有着近似的特性曲线，而我们也随机对几组数据进行了 origin 的图线拟合，得出了相似的结论。为了对于每一组数据都得到最优解，我们定义的评分函数如下：

$$s_{i,j} = \begin{cases} 0 & \text{if } |\hat{y}_{i,j} - y_{i,j}| \leq 1 \\ 1 & \text{if } 1 < |\hat{y}_{i,j} - y_{i,j}| \leq 2 \\ 2 & \text{if } 2 < |\hat{y}_{i,j} - y_{i,j}| \leq 3 \\ 4 & \text{if } 3 < |\hat{y}_{i,j} - y_{i,j}| \leq 5 \\ 10 & \text{if } |\hat{y}_{i,j} - y_{i,j}| > 5 \end{cases}$$

下标*i*代表样本序号  
下标*j*代表观测点序号  
对过大偏差的“惩罚性”

就是所谓的成本计算函数，本题要求的是我们需要得到成本的最小值。

我们的目标如下：

对某一样本的定标成本

$$S_i = \sum_{j=1}^{51} s_{i,j} + 20N_i$$

N<sub>i</sub>是该样本定标时所测定的点数  
第一项：误差成本  
第二项：测定成本

定标方案总成本

$$C = \frac{1}{M} \sum_{i=1}^M S_i$$

#### 3.2 多项式拟合

##### 2.1.1 几种多项式拟合的介绍

###### 1. 多项式拟合法

$$D = p_0 + p_1 * U + p_2 * U^2 + \dots + p_n * U^n \quad (2-1)$$

多项式的次数越高，曲线的拟合精度越高，但当次数过高时会导致有较大误差产生，因此，本次探讨采用三次多项式进行拟合

###### 2. 三参数幂指数法

$$D = D_0 + \beta U^\alpha \quad (2-2)$$

上式中，D<sub>0</sub>为初始值，α、β为待定常数

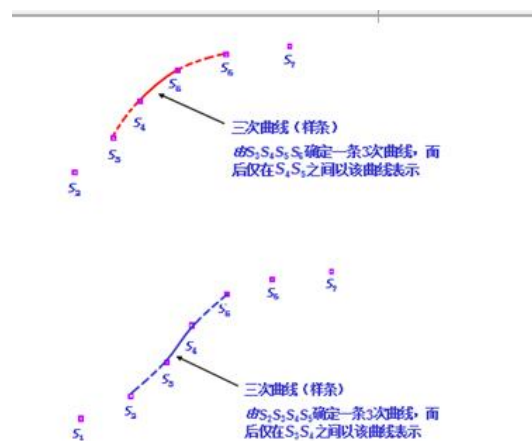
###### 3. e指数法

$$D = D_0 + B * e^{TU} \quad (2-3)$$

##### 2.1.2 结果分析

根据已有的数据进行分析，在多项式拟合曲线中，能够良好的反应出曲线的性状并保证运行速度的应为三、四、五次多项式，所以在寻找最优解的过程中，我们应当以上述三种多项式来进行拟合，从而达到最好的效果。然而同时多项式拟合也反映出了其精度不高的问题，即便是效果最好的三次、四次多项式，随机取点的样本中拟合出来并进行评分后所得到的结果也不是最理想，因此我们认为需要考虑更为精确的拟合手段。根据网络上的一些资源以及书籍资料，我们认为使用matlab自带样条插值应该会更精确。

样条插值是使用一种名为样条的特殊分段多项式进行插值的形式。由于样条插值可以使用低阶多项式样条实现较小的插值误差，这样就避免了使用高阶多项式所出现的龙格现象。



### 4 基于本课题的程序设计

#### 4.1 遗传算法基于 matlab 的实现

(1) 确定初始条件, 种群大小, 繁殖次数, 交叉互换概率, 突变概率, 随机生成五十一个个体作为初始种群, 此处采用 51 位的二进制数进行编码。

(2) 用 3.1 中所提到的评分函数作为适应度函数, 并分别使用三次样条差值法拟合曲线, 从而计算每个个体的适应度, 称之为 Q。

(3) 根据轮盘赌法, 使 Q 值小的拥有更多的繁殖机会, 通过交叉互换以及突变产生一个新种群。

(4) 记录每代中 Q 最小的个体，并继续进行下一代的繁殖，如此继续，直到达到设定的最大代数，此时记录中的最小个体即为我们得到的较优解。

### 4.2 设计中遇到的问题与分析

在设计底层函数的过程中，由于有网络资源的教学和参考，大部分都没有很大的问题，但是在设计适应度的时候，由于遗传算法是求解最大值，而本课题是求解正数最小值问题，因此我们选择将实际值取相反数并加上常数来运算，结束后再变换回来的方式，之后也尝试了直接取倒数的方式，效果在理论上是一样的。

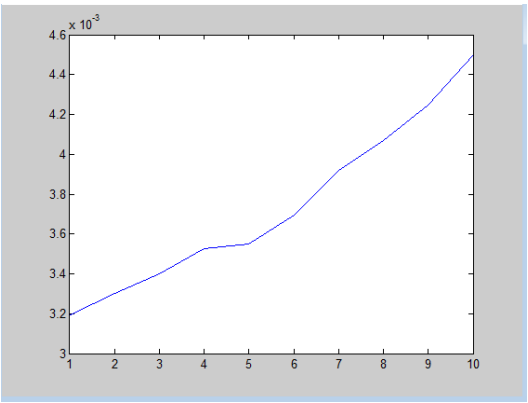
其次，我们的设计是对于每个点逐一计算，分析来说相当耗时，后来参考老师所给的测试文件，因为 matlab 对矩阵的计算有很大的优化，所以直接将源数据矩阵和拟合值矩阵相减之后直接计算会快很多，远比我们的设计有效率，这一点我们本应该在设计时就考虑到并优化。

## 5 实验测试数据及结果分析

### 5.1 遗传算法的初始条件设定

初步研究了网络的资源之后，我们开始设定初始条件，分别为：种群大小，染色体长度，遗传代数，交叉概率，变异概率。为了更好的解决这一类问题，我们将这些条件都交予用户自己设定，而本课题经过反复测试，我们制定的初始化条件分别为：种群大小 100，染色体长度 51，遗传代数 200，交叉概率 0.7，变异概率 0.015，并进行精英选择。交叉概率和变异概率是我们根据网上的建议数值进行了微调，而遗传代数则是通过自己的实验，实验过程是我们设定不同的初始值，利用控制变量法进行多次运行并归纳结果。

若取遗传代数为 10，则由下图可以看出，最终得出的最小成本值并不稳定并有下降趋势：



所以我们选择了 100 的数量级进行试验，最终得出在 80-100 代之间开始，最小成本趋于稳定。

### 5.2 最终取点结果以及最小成本值

经过之前的数据设置，我们进行了程序运行，200 代遗传所需运行时间大概为 60 分钟，符合之前实验的预期，总的来说还是十分耗时的。

在遗传至第 62 代时取到最优解。

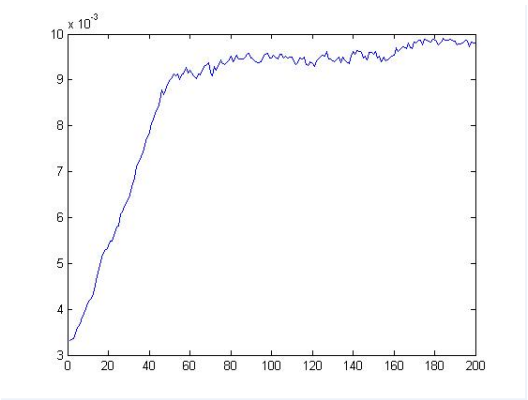
最终取点结果如下：

```
q =  
  
3      9      17      25      33      40      50
```

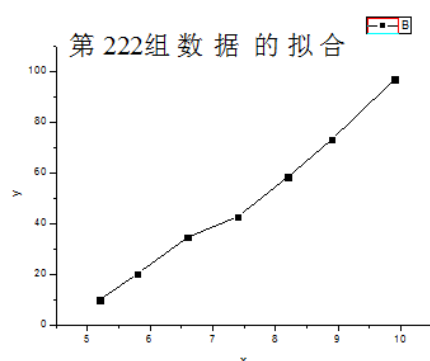
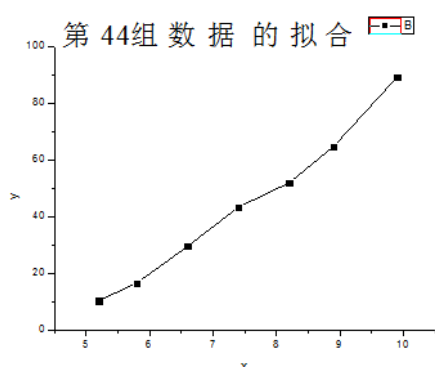
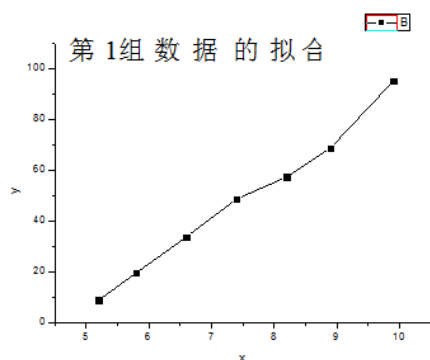
最小成本值如下：

```
best_fitness =  
|  
98.6930
```

对应的 (1/成本)-遗传代数函数图像如下：



由所得取点结果，我们随机选择了第 1、44、222 组数据进行了 origin 的图线处理拟合，得到的函数图像如下：



由上图可看出，图像的大致趋势与给出的预计效果接近。

## 6 结论和感想

### 6.1 结论

经过以上分析，我们得出遗传算法在该实际问题的统计推断应用中是可行的，而且是有效的，是有方向所取最优解的算法，可以很大程度上缩短计算的过程与时间，但其具体初始化的取值需要根据实际情况进行讨论。在三次插值法与三次多项式拟合法的比较中，我们得出三次插值法很大程度上优于三次多项式拟合法。

### 6.2 感想

经过这次课程的学习和实践，我们首先接触到了不同于 c++ 的编程语言 matlab，并在学习和实践代码的过程中加深了了解，但是由于编程质量和数量的不足，对于许多内置应用和函数的了解程度其实很低，造成实际效率的降低，这点希望能在今后对于 matlab 的使用中加大力度掌握。老师建议我们使用 matlab 来完成本课题一定有原因，相比于 c++，matlab 对于矩阵运算有很大的优化，因此非常适用于本课题，遗憾的是我们在适应度函数计算的时候并未很好的应用。

而在课程后期的师生面谈环节中，老师对于我们设计的误区以及不足有很好的讲解和引导，非常感谢老师当时的指引，避免了我们很多弯路，从而完成本课题。

## 7 参考文献

- [1]<http://baike.baidu.com/view/45853.htm> 百度百科
- [2]<http://blog.csdn.net/sunja7693/article/details/1408454> 网络资源，遗传算法的介绍与应用
- [3]<http://www.cnblogs.com/biaoyu/archive/2011/12/03/2274604.html> 网络资源，遗传算法的实现
- [4] 上海交通大学电子工程系, 第 3 讲: 问题的求解 (Part I) 基本任务框架下问题的求解, ppt 文件, 2010 年 10 月
- [5] MATLAB 数值计算: 机械工业出版社, 第 1 版 (2006 年 6 月 1 日)