一、案例名称

基于医药数据分析的曲线拟合算法及其实现

二、案例描述

一种新药进行试验并用于临床之前，需要设计一个合理的给药方案，确定每次给药的剂量和给药间隔时间等参数。药物进入机体后通过血液输送到全身，在这个过程中不断被吸收、分解、代谢，最终排除体外，药物在血液中的浓度即单位体积血液中药物含量，称为**血药浓度**。如果血药浓度太低，达不到预期的治疗效果；如果血药浓度太高，又可能导致药物中毒或副作用太大。获得给药后血药浓度随时间变化的规律是设计给药方案的基础。

根据药代动力学相关理论，在血液中，某些药物的血药浓度与时间的关系由下式描述：

 （2-1）

这里表示药物进入身体后的时间。这个模型的特点是当药物进入血液后药物浓度迅速上升，接着缓慢地指数型衰减。

氟西汀是一种常用的抗抑郁药，诺氟西汀（Norfluoxetine）是氟西汀的一个重要成分。当氟西汀被人体吸收后，会转化成诺氟西汀。为了更好地设计给药方案，已知病人血液中诺氟西汀的测量水平（单位：mg/ml即毫克每毫升）如表所示，请给出关系中的相关系数以确定药物浓度与时间的关系。

表1-1血液中药物浓度与时间的关系

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 小时 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 浓度（mg/ml） | 8.0 | 12.3 | 15.5 | 16.8 | 17.1 | 15.8 | 15.2 | 14.0 |

**输入**：血液中药物浓度与时间的关系测量水平数据。

**输出：**药物浓度与时间的关系由描述式：

 （2-2）

根据病人血液中已知药物诺氟西汀的测量水平数据，返回关系中的相关系数和。

#### **三、案例分析**

#### 1.曲线拟合的基本概念

为了得到药物浓度与时间的关系，我们需要确定公式（2-1）中的两个参数和，这本质上是一个曲线拟合的问题。下面我们先考察曲线拟合的基本概念。

假设通过实验或观测得到物理量与的一组离散的数据对其中其彼此不同。我们希望用较为简单的函数，如多项式函数，来反映物理量与之间的依赖关系。一般来说，当较大时，要求用多项式作为近似函数是不现实的。因为高次的插值多项式往往导致在非插值点处有较大的误差，同时实验和观测的数据对一般包含初始数据误差，应用严格通过点插值的办法也是不可取的。因此，我们希望寻找次数远比的数量低的多项式或其他简单函数，使它在一定度量意义下尽可能好地逼近或拟合原始的数据中的函数关系。

所谓曲线拟合问题，即是给定个数据对与某个函数集合(称为可取函数集合)，以及函数拟合于数据对的某种度量，需要在这个可取函数集合内寻找函数，使得这个度量变为最小。这里所说的拟合，即不要求所做的曲线完全通过所有的数据点，只要求所得的近似曲线能反映数据的基本趋势。

一个函数拟合于数据对的度量有许多种方式，最常见的一种是应用在各点处残差的平方和：

 （3-1）

选取这样的度量是为了避免偏差的正项与负项之间的抵消。由于我们取偏差的平方和作为函数拟合于数据对的度量，因此通常也称上述曲线拟合问题为**最小二乘拟合问题**。

#### 2.最小二乘法基本方法

为了找到拟合数据的函数，假设在平面坐标系上有这样一组点，我们希望在所有形如的直线中找到一条能使得直线到所有数据点的误差之和最小。这根误差之和最小的曲线我们称之为**拟合直线**。

如果采用一个带有高次项的多项式来代替这条曲线，能使得所有点上的误差之和最小，则这个表达式可以成为数据的多项式拟合函数。

一般地，给定数据，假设拟合函数的形式为：

 （3-2）

其中，为已知的线性无关函数。通过求系数，使得取最小值。

为了求的最小值点，令

 （3-3）

即

 （3-4）

由于是未知数，将上式整理为：

 （3-5）

该方程被称为**法方程**或者**正规方程**。这个方程的解可以确定拟合函数。

这种以残差平方和最小问题的解来确定拟合函数的方法称为**最小二乘法**。称为**变量之间的拟合函数**，也称为数据的**最小二乘拟合多项式。**

可以将拟合函数（3-1）写成两个向量的内积形式：

 （3-6）

其中是已经确定的一组函数。

对于多项式拟合问题，一般取基函数类，那么所求的拟合函数形式为，其中是待定系数。

把点代入，便得到以为未知数的方程组：

 （3-7）

由于方程组的个数多于未知数的个数，所以这类方程被称为**超定方程组**。其矩阵形式为：

 （3-8）

其中：



由此，在处的偏差就是超定方程组各方程的偏差。曲线拟合的条件就是确定，使得残差平方和达到最小值。于是，就是超定方程组的最小二乘解，根据公式（3-5）可知，求超定方程组的解其实也就是求法方程组的解。

#### 3.法方程组的求解

为了求得拟合函数的系数，需要求解法方程组的解。对于法方程组，可以采用如下方法来求解：



法方程组为：

 （3-9）

由上可知，对于一般的超定方程组，即：

 （3-10）

可用最小二乘法求解。

为了求解超定方程组，首先求出方程组系数矩阵的转置矩阵；然后计算矩阵和向量；最后求解法方程组法方程组的解称为原超定方程组(3-7)的最小二乘解。

超定方程组也称为**不相容方程组**。这个方程组无解。无解方程组可能是因为方程组的系数不精确引起的。在许多情形下，方程的个数大于未知数的个数，使得这些方程的解不能满足所有的方程。但是此时我们可以求一个接近于真实解的近似解。

于是，求超定方程组的过程可以用图3-1中的算法框图来描述：



图3-1 超定方程组的算法框图

**四、算法设计与分析**

解决了曲线拟合的一般方法问题，我们就可以把最小二乘法应用于诺氟西汀的药物浓度问题。此时，诺氟西汀的药物浓度与时间的关系实际上就转化为用最小二乘法求曲线拟合的问题。

在血液中，诺氟西汀的药物浓度与时间的关系由下式描述：

 （4-1）

这里表示药物引入身体后的时间。

为了方便采用最小二乘法，通过对二者取自然对数能把模型线性化得到：

 （4-2）

这里我们令。移项之后有：

 （4-3）

对于上述方程，药物浓度与时间都是已知数据，所求的未知参数为以及，要求出这两个未知参数，就是求矩阵方程的解，其中：

，

通过求解上述超定方程，可以求得以及，从而有。

在本题中，根据表1-1血液中药物浓度与时间的关系数据，

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 小时 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 浓度（mg/ml） | 8.0 | 12.3 | 15.5 | 16.8 | 17.1 | 15.8 | 15.2 | 14.0 |

可以得到：

， ，

其对应的法方程为：

 （4-4）

代入数据则有：

上述矩阵相乘得：

 （4-5）

求解上述法方程组，可得最小二乘解，即，那么，，则模型的拟合曲线是。

将最小二乘解代回公式中，可知最小二乘解的余项为：

 （4-6）

则，这也说明了上述方法的准确率较好。

1. **结语**

通过以上案例，我们可以看到，数学模型和计算机程序能够有效估计血液中药物浓度与时间的关系，从而为制定合理的给药方案提供理论依据。在模型的构建过程中，虽然做了一些必要的简化，但从拟合曲线和误差分析来看，整体效果比较理想。

然而，我们也要认识到，尽管数学模型和计算机模拟在药物研发和临床应用中具有重要作用，它们并不能完全替代临床试验和更深入的数据分析。我们应当保持严谨求实的科学态度，注重实践与理论的结合。

最重要的是，我们要始终树立“生命至上”的理念，将患者的健康放在首位，科学合理地指导用药。

1. **提供者和单位信息**

**教学案例提供者信息表**

|  |  |
| --- | --- |
| 案例名称 | 基于医药数据分析的曲线拟合算法及其实现 |
| 案例提供者 | 张老师、陈老师、刘老师  |
| 案例提供者单位 | 浙江工业大学计算机学院 |
| 案例联系人姓名 |   |
| 案例联系人电话 |   |
| 案例联系人邮箱 |   |