

数学建模思想方法大全及方法适用范围（上）

第一篇：方法适用范围

一、统计学方法

1.1 多元回归

1、方法概述：

在研究变量之间的相互影响关系模型时候，用到这类方法，具体地说：其可以定量地描述某一现象和某些因素之间的函数关系，将各变量的已知值带入回归方程可以求出因变量的估计值，从而可以进行预测等相关研究。

2、分类

分为两类：多元线性回归和非线性线性回归；其中非线性回归可以通过一定的变化转化为线性回归，比如： $y=\ln x$ 可以转化为 $y=u$ $u=\ln x$ 来解决；所以这里主要说明多元线性回归应该注意的问题。

3、注意事项

在做回归的时候，一定要注意两件事：

（1）回归方程的显著性检验（可以通过 sas 和 spss 来解决）

（2）回归系数的显著性检验（可以通过 sas 和 spss 来解决）

检验是很多学生在建模中不注意的地方，好的检验结果可以体现出你模型的优劣，是完整论文的体现，所以这点大家一定要注意。

4、使用步骤：

（1）根据已知条件的数据，通过预处理得出图像的大致趋势或者数据之间的大致关系；

（2）选取适当的回归方程；

（3）拟合回归参数；

（4）回归方程显著性检验及回归系数显著性检验

（5）进行后继研究（如：预测等）

1.2 聚类分析

1、方法概述

该方法说的通俗一点就是，将 n 个样本，通过适当的方法（选取方法很多，大家可以自行查找，可以在数据挖掘类的书籍中查找到，这里不再阐述）选取 m 聚类中心，通过研究各样本和各个聚类中心的距离 X_{ij} ，选择适当的聚类标准，通常利用最小距离法（一个样本归于一个类也就意味着，该样本距离该类对应的中心距离最近）来聚类，从而可以得到聚类结果，如果利用 sas 软件或者 spss 软件来做聚类分析，就可以得到相应的动态聚类图。

这种模型的特点是直观，容易理解。

2、分类

聚类有两种类型：

（1）Q 型聚类：即对样本聚类；

（2）R 型聚类：即对变量聚类；

通常聚类中衡量标准的选取有两种：

（1）相似系数法

（2）距离法

聚类方法：

（1）最短距离法（2）最长距离法

- (3) 中间距离法
- (4) 重心法
- (5) 类平均法
- (6) 可变类平均法
- (7) 可变法
- (8) 利差平均和法

在具体做题中，适当选区方法；

3、注意事项

在样本量比较大时，要得到聚类结果就显得不是很容易，这时需要根据背景知识和相关的其他方法辅助处理。

4、方法步骤

- (1) 首先把每个样本自成一类；
- (2) 选取适当的衡量标准，得到衡量矩阵，比如说：距离矩阵或相似性矩阵，找到矩阵中最小的元素，将该元素对应的两个类归为一类，
- (3) 重新计算类间距离，得到衡量矩阵
- (4) 重复第 2 步，直到只剩下一个类；

补充：聚类分析是一种无监督的分类，下面将介绍有监督的分类。

1.3 数据分类

1、方法概述

数据分类是一种典型的有监督的机器学习方法，其目的是从一组已知类别的数据中发现分类模型，以预测新数据的未知类别。这里需要说明的是：预测和分类是有区别的，预测是对数据的预测，而分类是类别的预测。

2、分类方法：

- (1) 神经网络
- (2) 决策树（这里不再阐述，有兴趣的同学，可以参考数据挖掘和数据仓库相关书籍）

3、注意事项

神经网络适用于下列情况的分类：

- (1) 数据量比较小，缺少足够的样本建立数学模型；
- (2) 数据的结构难以用传统的统计方法来描述
- (3) 分类模型难以表示为传统的统计模型

神经网络的优点：

分类准确度高，并行分布处理能力强，对噪声数据有较强的鲁棒性和容错能力，能够充分逼近复杂的非线性关系，具备联想记忆的功能等。

神经网络缺点：

需要大量的参数，不能观察中间学习过程，输出结果较难解释，会影响到结果的可信度，需要较长的学习时间，当数据量较大的时候，学习速度会制约其应用。

4、步骤

- (1) 初始化全系数
- (2) 输入训练样本
- (3) 计算实际输出值
- (4) 计算实际输出值和期望输出值之间的误差
- (5) 用误差去修改权系数

(6) 判断是否满足终止条件，如果满足终止，否则进入第二步

1.4 判别分析

1、概述

其是基于已知类别的训练样本，对未知类别的样本判别的一种统计方法，也是一种有监督的学习方法，是分类的一个子方法！

具体是：在研究已经过分类的样本基础上，根据某些判别分析方法建立判别式，然后对未知分类的样本进行分类！

2、分类

根据判别分析方法的不同，可分为下面几类：

- (1) 距离判别法
- (2) Fisher 判别法
- (3) Bayes 判别法
- (4) 逐步判别法

关于这几类的方法的介绍，大家可以参考《多元统计学》，其中比较常用的是 bayes 判别法和逐步判别法

3、注意事项：

判别分析主要针对的是有监督学习的分类问题。共有四种方法，这里重点注意其优缺点：

- (1) 距离判别方法简单容易理解，但是它将总体等概率看待，没有差异性；
- (2) Bayes 判别法有效地解决了距离判别法的不足，即：其考虑了先验概率——所以通常这种方法在实际中应用比较多！
- (3) 在进行判别分析之前，应首先检验各类均值是不是有差异（因为判别分析要求给定的样本数据必须有明显的差异），如果检验后某两个总体的差异不明显，应将这两个总体合为一个总体，再由剩下的互不相同的总体重现建立判别分析函数。
- (4) 这里说明下 Fisher 判别法和 bayes 判别法的使用要求：两者对总体的数据的分布要求不同，具体的，Fisher 要求对数据分布没有特殊要求，而 bayes 则要求数据分布是多元正态分布，但实际中却没有这么严格！
- (5) 可以利用 spss, sas 等软件来轻松实现

4、方法步骤

这里以 bayes 判别法为例讲述

- (1) 计算各类中变量的均值 \bar{x}_j 及均值向量 \bar{x}_h ，各变量的总均值 \bar{x}_i 及均值向量 \bar{x}
- (2) 计算类内协方差及其逆矩阵
- (3) 计算 bayes 判别函数中，各个变量的系数及常数项并写出判别函数
- (4) 计算类内协方差矩阵及各总协方差矩阵做多个变量的全体判别效果的检验
- (5) 做各个变量的判别能力检验
- (6) 判别样本应属于的类别

1.5 主成分分析

1、概述

主成分分析是一种降维数的数学方法，具体就是，通过降维技术立将多个变量量化为少数几个主成分的统计分析方法。在建模中，主要用于降维，系统评估，

回归分析，加权分析等等。

2、分类（无）

3、注意事项

在应用主成分分析时候，应该注意：

（1）综合指标彼此独立或者不想关

（2）每个综合指标所反映的各个样本的总信息量等于对应特征向量的特征值。

通常要选取的综合指标的特征值贡献率之和应为 80%以上

（3）其在应用上侧重于信息贡献影响力的综合评价

（4）当主成分因子负荷的符号有正也有负的时候，综合评价的函数意义就不明确！

4、方法步骤

大家可以参考《多元统计学》这本书籍，在这里就不做阐述

1.6 因子分析

1、概述

其是也是将变量总和为数量较少的几个因子，是降维的一种数学技术！它和主成分分析的最大区别是：其是一种探索性分析方法，即：通过用最少数量的几个不可观察的变量来说明出现在可观察变量中的相关模型（有点类似于前面讲述的分类和聚类的区别）它提供了一种有效的利用数学模型来解释事物之间的关系，体现出数据挖掘的一点精神！他可以用来解决降维，系统评估，加权分析等方法。

2、分类

因子分析是 R 型，即对变量研究

3、注意事项

（1）其不是对研究总体的变量的降维，而是根据原始变量信息构造新的变量，作为共同因子

（2）它通过旋转可以使得因子变量具有可解释性

（3）因子分析和主成分分析的区别和联系

<1>两者都是降维数学技术，前者是后者的推广和发展

<2>主成分分析只是一般的变量替换，其始终是基于原始变量研究

数据的模型规律；而因子分析则是通过挖掘出新的少数变量，来研究的一种方法，有点像数据挖掘中的关联规则发现！

4、方法步骤

（略）大家可以去论坛上下载相关电子资源，也可以参考《多元统计学》

1.7 残差分析

1、概述

在实际问题中，由于观察人员的粗心或偶然因素的干扰。常会使我们所得到的数据不完全可靠，即出现异常数据。有时即使通过相关系数或 F 检验证实回归方程可靠，也不能排除数据存在上述问题。残差分析的目的就在于解决这一问题。所谓残差是指实际观察值与回归估计值的差。

2、分类

无

3、应用

（1）通过残差分析来排除异常数据

(2) 通过残差分析来检验模型的可靠性

还有很多应用，大家在使用过程中据情况选取，灵活应用！

1.8 典型相关分析

1、概述

前面介绍的方法主要是一个变量和多个变量之间的关系，而典型相关分析研究的是多个变量和多个变量之间的关系，或者是一组变量和一组变量之间关系！其可以揭示两组变量之间的关系，从而供大家研究两个现象之间的关系，例如：蔬菜的产出水平和影响产出水平的变量之间的关系！

2、分类

多对多的变量关系研究！

3、注意事项

(1) 其可以很好地解决组合相关性的问题

(2) 其还局限于两组变量的研究，而且要求这两组变量都是连续变量且需服从多元正态分布

1.9 时间序列

1、概述

时间序列预测法是一种定量分析方法，它是在时间序列变量分析的基础上，运用一定的数学方法建立预测模型，使时间趋势向外延伸，从而预测未来市场的发展变化趋势，确定变量预测值。

其基本特点是：假定事物的过去趋势会延伸到未来；预测所依据的数据具有不规则性；撇开市场发展之间的因果关系。

2、分类

时间序列的变动形态一般分为四种：长期趋势变动，季节变动，循环变动，不规则变动。

方法分类：

- (1) 平均数预测（简单算术平均法，加权算术平均法，几何平均数法）
- (2) 移动平均数预测（一次移动平均法，二次移动平均法）
- (3) 指数平滑法预测（一次，二次，三次指数平滑法）
- (4) 趋势法预测（分割平均法，最小二乘法，三点法）
- (5) 季节变动法（简单平均法，季节比例法）

3. 注意事项

(1) 季节变动法预测需要筹集至少三年以上的资料

(2) 移动平均法在短期预测中较准确，长期预测中效果较差；

(3) 移动平均可以消除或减少时间序列数据受偶然性因素干扰而产生的随机变动影响。

(4) 一次移动平均法适用于具有明显线性趋势的时间序列数据的预测；一次移动平均法只能用来对下一期进行预测，不能用于长期预测，必须选择合理的移动跨期，跨期越大对预测的平滑影响也越大，移动平均数滞后于实际数据的偏差也越大。跨期太小则又不能有效消除偶然因素的影响。跨期取值可在 3~20 间选取。

(5) 二次移动平均法与一次移动平均法相比，其优点是大大减少了滞后偏差，使预测准确性提高；二次移动平均只适用于短期预测。而且只用于的情形。

(6) 最小二乘法即适用于直线趋势的预测，也适用于曲线趋势的预测。

还有一些注意事项，这里就不在意义罗列

4. 方法步骤

(略)

统计学的分析方法到这先告一段落！

数学建模思想方法大全及方法适用范围（中）

优化方法

二、优化方法

1、概述

在一系列的条件限制下，寻求最优方案，使得目标达到最优的问题统称为优化问题。解决这类问题的方法，自然就称之为优化方法，又成为数学规划！其是运筹学的一个重要分支！

2、分类

优化问题可以归结为优化模型，按照优化模型求解方法的不同，可以分为以下类别：

(1) 按照有无约束条件：无约束和约束最优化问题

(2) 按照决策变量是否取之连续分为：

a) 数学规划或连续规划：LP, NLP, QP

b) 离散优化或组合优化：IP

(3) 单目标规划和多目标规划

(4) 确定性规划和不确定性规划

(5) 目标规划，动态规划，非线性规划，多目标规划等

3、注意事项

(1) 约束优化问题可以转化为无约束优化问题来解决

(2) 多目标规划可以通过适当的方法转化为单目标规划来解决

(3) 非线性规划，在一定的条件下，可以近似为线性规划来解决

(4) 不确定性规划可以通过适当的技巧转化为确定性方法解决这些转化，

大家可以参阅优秀论文，注意学取人家的经验，从而建立自己的能力提升！

也可以参考相关优化书籍和运筹学书籍来学习，书上都有介绍！

4、步骤

简要概括

(1) 建立规划模型

(2) 选择方法

(3) 结果解释

三、排队论

1. 概述

在我们的生活中，经常会做和排队相关的事情，比如：银行等待取钱，医院挂号排队，理发排队等等，都会涉及到排队问题，并且 2009 年国赛 B 题第五小题就考了和排队论相关的问题。

排队论是一门研究拥挤现象的学科，具体就是研究各种排队系统概率基础

上，解决相应的排队系统的最优设计和最优控制问题（注意：其实解决排队最优设计和最优控制的问题）

2、分类

(1) 泊松输入-指数服务排队模型

(2) M/M/1 等待排队模型

(3) M/M/S/inf 模型 (2009 年 B 题的排队就是一个这种模型)

3、注意事项

(1) 排队系统常见的优化问题在于

- a) 确定最优服务率
- b) 确定最佳服务台数量
- c) 选择最为合适的服务规则
- d) 确定上述几个量的最优组合

(2) 在使用时候, 应该注意模型的选择, 模型选择好后, 注意顾客到达和服务时间分布的选择和检验

4、步骤

(略)

四：智能算法的优化应用

遗传算法, 蚁群算法, 贪婪算法, 模拟退火算法, 回溯法, 分支定界法, 禁忌搜索算法, 粒子群算法等启发式算法。

这类算法在实际中应谨慎使用, 固然这类算法有其好处, 但是其不确定远远大于其确定性, 所以大家在比赛中慎重使用, 最好使用成熟的算法!

五、微分建模

1、使用背景介绍

我们通常会在研究一些问题时候, 涉及到某些变量的变化率或导数, 这样所得到的变量之间的关系就是微分方程模型, 其反映的是变量之间的间接关系, 通过求解方程, 就可以得到直接关系!

2、分类

按照求解方法的不同分为: 精确求解, 求数值解, 定性分析方法

3、注意事项

(1) 微分方程的集中建立方法:

- A、根据规律建立方程
- B、微元法
- C、模拟近似

(2) 几种常见的微分方程模型:

人口增长模型, 传染病模型, 捕鱼模型等

4、步骤

(略)

六、差分方程

1、使用背景的介绍

它是研究离散变量的变化规律, 具体就是: 根据实际的规律性质, 平衡关系等, 建立离散变量所满足的平衡关系式, 从而建立差分方程模型。通过求出和分析解, 研究方程解的性质, 比如, 平稳性, 渐进性, 振动性, 周期性等等

2、分类

几个重要的差分方程模型:

商业贷款, 离散形式的人口模型, 汽车租赁, 动物养殖问题等经典模型, 大家可以参考, 然后对差分方程模型有个全面的认识

七、图论

由于图论设计的内容比较多, 系统庞杂, 而且模型比较灵活, 所以在这里不做介绍, 大家可以自行研究。

八、其他方法介绍

灰色系统主要用来解决少数据的预测问题；层次分析法主要用来解决综合分析类的问题；模糊数学可以用来做模糊综合评判，模糊聚类分析和模糊线性规划；蒙特卡洛方法主要是一种计算机仿真方法，通常在排队论问题使用；神经网络可以用来分类，预测，建立模型等等；还有一些其他的方法，这里就不一一介绍了。这里我再补充一点：

预测方法的比较：

一般小样本内部预测用插值和拟合，大样本内部预测用回归模型。

移动平均法属于时间序列分析的内容，时间序列分析一般用作大样本外部预测，就是对未来的预测，灰色理论用于小样本的外部预测，神经网络用于超大样本的未来预测。

可以把这些预测方法归纳一下，对于建模，主要考虑这些方法的适用性。

上面主要是对《建模方法的使用范围》做了一定的介绍，关于《数学建模中的思想方法大全》

数学建模思想方法大全及方法适用范围（下）

第二篇《数学建模思想方法大全》

在数学建模中常用的方法：类比法、二分法、量纲分析法、差分法、变分法、图论法、层次分析法、数据拟合法、回归分析法、数学规划（线性规划，非线性规划，整数规划，动态规划，目标规划）、机理分析、排队方法、对策方法、决策方法、模糊评判方法、时间序列方法、灰色理论方法、现代优化算法（禁忌搜索算法，模拟退火算法，遗传算法，神经网络）。

用这些方法可以解下列一些模型：优化模型、微分方程模型、统计模型、概率模型、图论模型、决策模型。

拟合与插值方法（给出一批数据点，确定满足特定要求的曲线或者曲面，从而反映对象整体的变化趋势）：matlab 可以实现一元函数，包括多项式和非线性函数的拟合以及多元函数的拟合，即回归分析，从而确定函数； 同时也可以利用 matlab 实现分段线性、多项式、样条以及多维插值。

在优化方法中，决策变量、目标函数（尽量简单、光滑）、约束条件、求解方法是四个关键因素。其中包括无约束规则（用 fminsearch、fminbnd 实现）线性规则（用 linprog 实现）非线性规则、（用 fmincon 实现）多目标规划（有目标加权、效用函数）动态规划（倒向和正向）整数规划。

回归分析：对具有相关关系的现象，根据其关系形态，选择一个合适的数学模型，用来近似地表示变量间的平均变化关系的一种统计方法（一元线性回归、多元线性回归、非线性回归），回归分析在一组数据的基础上研究这样几个问题：建立因变量与自变量之间的回归模型（经验公式）；对回归模型的可信度进行检验；判断每个自变量对因变量的影响是否显著；判断回归模型是否适合这组数据；利用回归模型对进行预报或控制。相对应的有线性回归、多元二项式回归、非线性回归。

逐步回归分析：从一个自变量开始，视自变量作用的显著程度，从大到地依次逐个引入回归方程；当引入的自变量由于后面变量的引入而变得不显著时，要将其剔除掉；引入一个自变量或从回归方程中剔除一个自变量，为逐步回归的一步；对于每一步都要进行值检验，以确保每次引入新的显著性变量前回归方程中只包含对作用显著的变量；这个过程反复进行，直至既无不显著的变量从回归方

程中剔除，又无显著变量可引入回归方程时为止。（主要用 SASmatlab matlab 来实现，也可以用 mmaattllaabb 软件来实现）。

聚类分析：所研究的样本或者变量之间存在程度不同的相似性，要求设法找出一些能够度量它们之间相似程度的统计量作为分类的依据，再利用这些量将样本或者变量进行分类。

系统聚类分析：将 n 个样本或者 n 个指标看成 n 类，一类包括一个样本或者指标，然后将性、质最接近的两类合并成为一个新类，依此类推。最终可以按照需要来决定分多少类，每类有多少样本（指标）。

系统聚类方法步骤：

1. 计算 n 个样本两两之间的距离
2. 构成 n 个类，每类只包含一个样品
3. 合并距离最近的两类为一个新类
4. 计算新类与当前各类的距离（新类与当前类的距离等于当前类与组合类中包含的类的距离最小值），若类的个数等于 1，转 5，否则转 3
5. 画聚类图
6. 决定类的个数和类。

判别分析：在已知研究对象分成若干类型，并已取得各种类型的一批已知样品的观测数据，在此基础上根据某些准则建立判别式，然后对未知类型的样品进行判别分类。

距离判别法—首先根据已知分类的数据，分别计算各类的重心，计算新个体到每类的距离，确定最短的距离（欧氏距离、马氏距离）

Fisher 判别法—利用已知类别个体的指标构造判别式（同类差别较小、不同类差别较大），按照判别式的值判断新个体的类别

Bayes 判别法—计算新给样品属于各总体的条件概率，比较概率的大小，然后将新样品判归为来自概率最大的总体

模糊数学：研究和处理模糊性现象的数学（概念与其对立面之间没有一条明确的分界线）与模糊数学相关的问题：模糊分类问题—已知若干个相互之间不分明的模糊概念，需要判断某个确定事物用哪一个模糊概念来反映更合理准确；模糊相似选择—按某种性质对一组事物或对象排序是一类常见的问题，但是用来比较的性质具有边界不分明的模糊性；模糊聚类分析—根据研究对象本身的属性构造模糊矩阵，在此基础上根据一定的隶属度来确定其分类关系；模糊层次分析法—两两比较指标的确定；模糊综合评判—综合评判就是对受到多个因素制约的事物或对象作出一个总的评价，如产品质量评定、科技成果鉴定、某种作物种植适应性的评价等，都属于综合评判问题。由于从多方面对事物进行评价难免带有模糊性和主观性，采用模糊数学的方法进行综合评判将使结果尽量客观从而取得更好的实际效果。

时间序列是按时间顺序排列的、随时间变化且相互关联的数据序列—通过对预测目标自身时间序列的处理，来研究其变化趋势（长期趋势变动、季节变动、循环变动、不规则变动）

自回归模型：一般自回归模型 $AR(n)$ —系统在时刻 t 的响应 $X(t)$ 仅与其以前时刻的响应 $X(t-1), \dots, X(t-n)$ 有关，而与其以前时刻进入系统的扰动无关；移动平均模型 $MA(m)$ —系统在时刻 t 的响应 $X(t)$ ，与其以前任何时刻的响应无关，而与其以前时刻进入系统的扰动 $a(t-1), \dots, a(t-m)$ 存在着一定的相关关系；自回归移动平均模型 $ARMA(n, m)$ —系统在时刻 t 的响应 $X(t)$ ，不仅与其前 n 个时刻的自

身值有关，而且还与其前 m 个时刻进入系统的扰动存在一定的依存关系。

时间序列建模的基本步骤

1. 数据的预处理：数据的剔取及提取趋势项
2. 取 $n=1$ ，拟合 $ARMA(2n, 2n-1)$ （即 $ARMA(2, 1)$ ）模型
3. $n=n+1$ ，拟合 $ARMA(2n, 2n-1)$ 模型
4. 用 F 准则检验模型的适用性。若检验显著，则转入第 2 步。若检验不显著，转入第 5 步。
5. 检查远端时刻的系数值的值是否很小，其置信区间是否包含零。若不是，则适用的模型就是 $ARMA(2n, 2n-1)$ 。若很小，且其置信区间包含零，则拟合 $ARMA(2n-1, 2n-2)$ 。
6. 利用 F 准则检验模型 $ARMA(2n, 2n-1)$ 和 $ARMA(2n-1, 2n-2)$ ，若 F 值不显著，转入第 7 步；若 F 值显著，转入第 8 步。
7. 舍弃小的 MA 参数，拟合 $m < 2n-2$ 的模型 $ARMA(2n-1, m)$ ，并用 F 准则进行检验。重复这一过程，直到得出具有最小参数的适用模型为止
8. 舍弃小的 MA 参数，拟合 $m < 2n-1$ 的模型 $ARMA(2n, m)$ ，并用 F 准则进行检验。重复这一过程，直到得出具有最小参数的适用模型为止。

图论方法：

最短路问题：两个指定顶点之间的最短路径—给出了一个连接若干个城镇的铁路网络，在这个网络的两个指定城镇间，找一条最短铁路线（Dijkstra 算法）
每对顶点之间的最短路径（Dijkstra 算法、Floyd 算法）。

最小生成树问题：连线问题—欲修筑连接多个城市的铁路设计一个线路图，使总造价最低（prim 算法、Kruskal 算法）。

图的匹配问题：人员分派问题： n 个工作人员去做件 n 份工作，每人适合做其中一件或几件，问能否每人都有一份适合的工作？如果不能，最多几人可以有适合的工作？（匈牙利算法）。

遍历性问题：中国邮递员问题—邮递员发送邮件时，要从邮局出发，经过他投递范围内的每条街道至少一次，然后返回邮局，但邮递员希望选择一条行程最短的路线

最大流问题，运输问题：

最小费用最大流问题：在运输问题中，人们总是希望在完成运输任务的同时，寻求一个使总的运输费用最小的运输方案

在数学建模中常用的算法：

- 1: 蒙特卡罗算法；
- 2: 数据拟合、参数估计、插值等数据处理算法（常用 matlab 实现）；
- 3: 线性规划、整数规划、多元规划、二次规划(用 lingo、lingdo、matlab 即可实现)；
- 4: 图论算法（包括最短路、网络流、二分图）；
- 5: 动态规划、回溯搜索、分治算法、分支界定；
- 6: 最优化理论的三大经典算法（模拟退火算法、神经网络算法、遗传算法）；
- 7: 网格算法和穷举法；
- 8: 连续数据离散化；
- 9: 数值分析算法；
- 10: 图象处理算法（常用 matlab 来实现）。

