保奖课程算法汇总 by wsx

数据处理方法：

一、插值与拟合问题

1.概念

在实际中，常常要处理由实验或测量所得到的一些离散数据。插值与你和方法就是要通过这些数据去确定某一类已知函数的参数或寻求某个近似函数，使所得到的近似函数与已知数据有较高的拟合精度。

如果要求这个近似函数（曲线或曲面）经过所已知的所有数据点，则称此类问题为插值问题。

如果不要求近似函数通过所有数据点，而是要求它能较好地反映数据变化规律的近似函数的方法称为数据拟合。

2.插值与拟合的区别与联系

（1）联系

根据实际中一组已知数据来构造一个能够反映数据变化规律的近似函数的方法。

（2）区别

插值问题不一定得到近似函数的表达形式，仅通过插值方法找到未知点对应的值。数据拟合要求得到一个具体的近似函数的表达式。

3.matlab实现插值

```
yi=interp1(x,y,xi,'method');
%yi为xi处的插值结果，x、y为插值节点，xi为被插值点，method为插值方法
%method有以下几种
%'nearest' 最邻近插值
%'linear' 线性插值
%'spline' 三次样条插值
%'cubic' 立方插值
%缺省时，分段线性插值
%e.g
hours=1:12;
temps=[5,8,9,15,25,29,31,30,22,25,27,24];
h=1:0.1:12;
t=interp1(hours,temps,h,'spline');
plot(hours,temps,'+',h,t,hours,temps,'r:');
z=interp2(x0,y0,z0,x,y,'method');
%x0、y0、z0为插值节点，x、y为被插值点
%'nearest' 最邻近插值
%'linear' 双线性插值
%'cubic' 双三次插值
%缺省时，双线性插值
```

4.matlab实现拟合

```
%线性拟合：
a=ployfit(x,y,m);
%x和y为已知点的坐标，一一对应 m为拟合多项式的次数
y=ployval(a,x);
%拟合好之后计算x坐标下的y值
%指数增长模型拟合
%y=ae^(bx)
y=ployfit(x,log(y),1);
```

二、数学算法

1.常微分方程

2.线性方程与常数变易法

3.恰当方程与积应分子

4.一阶隐式微分方程及其参数表示

评价类题目以及分类方法：

一、层次分析法

1.概念

层次分析法是将定性问题定量化处理的一种有效手段。主要有机理分析法和统计分析法两种方法。

2.基本步骤

1.建立层次结构模型 一般分为目标层、方案层、准则层

买钢笔——目标层

质量、颜色、价格、外形、实用——准则层

钢笔1、钢笔2、钢笔3、钢笔4——方案层

（1）质量、颜色、价格、外形、实用进行排序

（2）将各个钢笔的质量、颜色、价格、外形、实用进行排序

（3）经综合分析决定买哪支钢笔。与人们对某一复杂决策问题的思维、判断过程大体一致。

2.构造成对比较阵

每次取两个因素xi和xj，用正数aij表示xi与xj的重要性之比。

3.检验矩阵A的一致性

二、TOPSIS算法的原理与实现

1.概念

优劣解距法，该方法能够根据现有的数据，对个体进行评价排序，根据有限个评价对象与理想化目标的接近程度进行排序的方法，是在现有的对象中进行相对优劣的评价。

2.例子

小明分数90 80 小华75 95 小亮80 80 问谁的综合能力最好

正理想解D+:{100,100}

负理想解D-：{60,60}

采用绝对值距离或欧氏距离

D+=|100-90|+|100-80|=30

D-=|60-90|+|60-80|=50

评价指标系数C=D-/(D++D-)=5/8

3.指标的处理

（1）极大型指标 成绩、利润

（2） 极小型指标 费用、废品率

（3） 中间型 水质PH值、体温

处理方法：图（1）

三、综合评价法

1.概念

综合评价指运用多个指标对多个参评单位进行评价的方法，称为多变量综合评价方法，又称综合评价法，其基本思想是将多个指标转化为一个能够反映综合情况的指标来进行评价。

感觉就是将一个待评价物体的各个属性乘上一个权重，最后根据结果对这些物体进行一个分类或者排序。

e.g. 小麦倒伏、空气污染

特点：

（1）多个指标评价同时完成

（2）根据指标的重要性加权处理

（3）评价结果不再是具有具体含义的统计指标，而是以指数或分值表示参评单位“综合状况”的排序

一般表现为以下几个问题：

（1）分类——对所研究对象的全部个体进行分类

（2）比较、排序

（3）考察某一综合指标的整体实现程度，如小康目标的实现程度、现代化的实现程度

五个要素：被评价对象、评价指标、权重系数、综合评价模型和评价者

综合评价的一般步骤

1.确定目的

2.建立评价指标体系

3.对指标数据做预处理

4.确定各个指标的权重 主管定权法、客观定权法

2.常用综合评价方法

（1）线性加权综合法 就是权值乘上属性值，非常简单 适用于各评价指标之间相互独立，但对于不完全独立的情况，其结果将导致各指标间信息的重复

（2）非线性加权综合法，用一个非线性函数作为综合评价模型，适用于各指标之间具有较强关联性 对数据要求较高，指标数值不能为0、负数 且乘法易拉开评价档次

（3）TOPSIS，前面提到了

（4）灰色关联分析法，下面提到

3.模糊综合评价

对方案、人才、成果的评价，人们考虑的因素很多，且有些描述很难给出确切表达

e.g.

对某电视机进行综合模糊评价

（1）确定指标集U={图像、声音、价格}和评语集{很好、较好、一般、不好}

（2）求解模糊评价矩阵，例如对于图像有30%认为很好，50%认为较好，20%认为一般，0%认为不好，然后把这些作为矩阵的一行，另外两行同理，设定三个指标的权系数向量，A=(0.5,0.3,0.2)这个系数可以用层次分析法来定，就是把两个指标两两比较

（3）把评价矩阵和权系数矩阵相乘，再归一化处理B=AP,B=(0.25,0.42,0.17,0.17)则电视机比较好。

四、灰色系统理论

1.应用

（1）灰色关联分析 对于某一个总值来说，若一个分值和大值关联较大，则分值的变化会和大值息息相关，这个分析法可以算出分值对总值的关联系数 e.g.给出某地连续七年的总收入、养猪业和养兔业数值，画图发现养猪业的曲线变化趋势和总收入曲线变化趋势基本相同，则说明养猪业的关联系数较大

（2）灰色预测：人口预测、初霜预测、灾变预测

（3）灰色决策

（4）灰色预测控制

五、决策树

1.概念

决策树是一种十分常用的分类方法，需要监管学习，监管学习就是给出一堆样本，每个样本都有一组属性和一个分类结果，通过学习这些样本得到一个决策树，这个决策树能够对新的数据给出正确的分类。

决策树是一种树形结构，其中每个内部节点表示一个属性上的判断，每个分支代表一个判断结果的输出，最后每个叶节点代表一种分类结果。

2.决策树的生成

（1） 节点的分裂：一般当一个节点所代表的属性无法给出判断时，则选择将这一节点分成2个

子节点（如不是二叉树的情况会分成n个子节点）

（2）阈值的确定：选择适当的阈值使得分类错误率最小。

3.常见决策树

（1）ID3-“最大化信息增益”

（2）C4.5-“信息增益率”

（3）CART-“既可以做分类也可以做回归”

六、随机森林

1.简介

随机森林的基本单元就是决策树，如果我们需要将一个输入样本进行分类，那么就小将它输入到每棵树中进行分类，将若干个弱分类器的分类结果进行投票选择，从而组成一个强分类器。

2.特点

（1）能够处理具有高维特征的输入样本，而且不需要降维

（2）能够评估各个特征在分类问题上的重要性

（3）在生成过程中，能够获取到内部生成误差的一种无偏估计

（4）对于缺省值问题也能够获得很好得结果

3.实现

应该可以调佣函数库来实现，我看有一个库叫sklearn，其中有RandomForestClassifier方法。

七、主成分分析法

1.主要作用

将多个指标进行降维处理，变成少数几个重要指标，看上去原理是用矩阵相乘，把多个指标按权重相乘，形成一个新的指标，我们可以再对这个新的指标赋予意义。

2.主要步骤

（1）对原始数据进行标准化处理

（2）计算相关系数矩阵R

（3）计算矩阵R的特征值和特征向量

（4）选择p个主成分

八、神经网络

我觉得DNN、CNN大家都比较熟，也知道是干啥的，不多写。

求最值问题：

一、二次规划

1.定义

二次规划在运筹学中是一种特殊类型的最佳化的问题，是解决特殊类型的数学优化问题的过程，是一个线性约束的二次优化问题，即优化几个受线性变量影响的二次函数的问题对这些变量的限制，二次规划是一种特殊类型的非线性规划

非线性规划：求解最优化问题，伴随着一个要被最大化或最小化的目标函数，知识一些约束或目标函数是非线性的，它是最优化处理非线性问题的一个子领域。

这个因为数学知识基本忘了所以没太理解，需要用到凸函数、凹函数、核心是求Hessian矩阵，求完以后代到一个式子里就行。从图上看是求一个函数的全局最优解、局部最优解等等。如果题目要在一些线性或者非线性约束条件下求最值，并且把函数建模出来了可以细细研究这个方法。

2.案例举例

假设有四种投资：社保债券、技术交易中心、管理咨询中心、游乐中心，令第i种投资的收益率ri的方差表示投资的奉贤大小，即收益率关于均值的偏离程度，令xi为第i个项目的投资额占总投资的比例，向量x=（x1，x2，x3，x4)表示一个投资组合，则其对应的收益率为R=rx，预期收益不低于8.5。

3.matlab求解

Fmincon函数是matlab最主要内置的求解约束最优化的函数

二、图论算法

这个数据结构课上有讲，就是DIJKSTRA、Floyd这些东西，估计ACM校队老哥也懂所以也不多写。

三、遗传算法

1.定义

遗传算法是模拟达尔文生物进化论的自然选择和遗传学机理的生物进化过程的计算模型，是一种通过模拟你自然进化过程搜索最优解的方法。

主要特点：

（1）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

（2）具有内在的隐并行性和更好的全局寻优能力

（3）采用该氯化的寻优方法

2.找最优值的算法，用袋鼠跳来理解

（1）爬山算法

一只袋鼠朝着比现在高的地方跳去。它找到了不远处的最高的山峰。但是这座山不一定是最高峰。这就是爬山算法，它不能保证局部最优值就是全局最优值。

（2）模拟退火，下面会讲

袋鼠喝醉了。它随机地跳了很长时间。这期间，它可能走向高处，也可能踏入平地。但是，它渐渐清醒了并朝最高峰跳去。这就是模拟退火算法。

（3）遗传算法

有很多袋鼠，它们降落到喜玛拉雅山脉的任意地方。这些袋鼠并不知道它们的任务是寻找珠穆朗玛峰。但每过几年，就在一些海拔高度较低的地方射杀一些袋鼠。于是，不断有袋鼠死于海拔较低的地方，而越是在海拔高的袋鼠越是能活得更久，也越有机会生儿育女。就这样经过许多年，这些袋鼠们竟然都不自觉地聚拢到了一个个的山峰上，可是在所有的袋鼠中，只有聚拢到珠穆朗玛峰的袋鼠被带回了美丽的澳洲。

3.一般步骤

（1）随机产生种群。

（2）根据策略判断个体的适应度，是否符合优化准则，若符合，输出最佳个体及其最优解，结束。否则，进行下一步。

（3）依据适应度选择父母，适应度高的个体被选中的概率高，适应度低的个体被淘汰。

（4）用父母的染色体按照一定的方法进行交叉，生成子代。

（5）对子代染色体进行变异。

（6）由交叉和变异产生新一代种群，返回步骤2，直到最优解产生。

四、多目标规划

1.适用范围

研究多余一个的目标函数在给定区域上的最优化。

在很多问题中，例如经济、管理、军事、科学和工程设计等领域，衡量一个方案的好坏往往难以用一个指标来判断，而需要用多个目标来比较，这些目标有时不甚协调，甚至是矛盾的，因此有许多学者致力于这方面的研究。

2.与线性规划的区别

（1）线性规划只讨论一个线性目标函数在一组线性约束条件下的极值问题；目标规划是多个目标决策，可求得更切合实际的解。

（2）线性规划要求问题的解必须严格满足全部约束条件，目标规划无此要求，实际问题中并非所有约束都需严格满足。

（3）线性规划求最优解，目标规划是找满意解。

（4）线性规划中的约束条件是同等重要的，是硬约束；目标规划中有轻重缓急和主次之分，即有优先权。

五、粒子群算法

1.概念

模拟鸟群的捕食行为，一群鸟在随机搜索食物，在这个区域里只有一块食物，所有的鸟都不知道食物在哪里，但是它们都知道当前位置离食物有多远，那么找到食物的最优策略是搜寻离食物最近的鸟的周围区域。

2.适用举例

假设要找一个函数z=f（x,y)的最大值，首先会生成一些随机的x和y值，然后根据这些x和y计算出z，在所有的z里找一个最大值z0，那么真正的最大值可能就在这个最大值附近，于是对所有随机生成的点进行修正，让它们靠近与z0对应的x和y。

六、模拟退火算法

1.概念

模拟将固体加热到一定温度，保持足够时间，然后以适宜速度冷却，加温时，固体内部粒子随温升变为无序状，内能增大，徐徐冷却时粒子渐趋有序，在每个温度都到达平衡态，然后在常温时达到基态，内能减为最小。

从直观函数图来看，一个函数有一些山峰和山谷，模拟退火算法当温度较高时跳跃幅度会比较大，这就让它有概率能够跳出局部最优的范围，去找到全局最优解。当温度降低以后跳跃的幅度逐渐减小，这就方便去求出最优值。

七、差分进化算法

1.概念

和遗传算法差不多，只不过遗传算法是先进行选择，选择操作是轮盘赌法，把每个个体的适应度除以总适应度，适应度越大的因子越容易被选中遗传，然后进行交叉操作，随机选择两个因子，将某一个位点交叉互换，最后进行变异操作，即随机选择一个个体，再随机选择一个位点，再按照公式进行变异。

差分进化法是先进行变异，后进行交叉，最后进行选择。

2.适用范围

遗传算法的变异可能会与原先种群中某个体的基因重合，这种变异是没有意义的，因为并没有产生新的解，后果是到了优化的后期，整个种群可能陷入局部最优。差分算法构造出一些与众不同的变异，也就是和当前的所有解区分开来，从而有可能碰上全局最优。

预测类题目：

一、回归分析算法

1.概念

基础物理实验也用过，我觉得意思就是设几个未知参数，然后根据已有的点在误差允许范围内利用最小二乘法去求参数的值。

2.一元线性回归

Y=β0+β1x，称为y对x的回归直线方程

作用：

（1）对Y做预测

（2）对Y的范围做控制

主要任务：

（1）用试验值对β0、β1作点估计

*（2）对回归系数β0、β1作假设检验

（3）在x=x0处对y作预测，对y作区间估计

3.多元线性回归

把X、Y换成矩阵，仍用最小二乘法的思想

4.非线性回归

y=ae^bx

5.适用范围

样本数量较少，自变量与因变量之间的变化具有明显的逻辑关系

二、马尔可夫模型

1.适用范围

这个马尔可夫模型应该是预测未来某件事发生的概率，且是由现在的状态去预测下一个状态的概率。最后有一个矩阵表示下一时刻各个状态的概率大小。

2.随机过程定义

设{epsilont，t∈T}是一族随机变量，T是一个实数集合，若对任意实数t∈T，epsilont是一个随机变量，则称{epsilont，t∈T}为随机过程

3.适用范围

马尔可夫链是一类特殊的随机序列，马尔可夫模型指的是某一系统在已知现在的情况下，系统未来时刻的情况只与现在有关，而与过去的历史无直接关系，例如投篮。

主要用于市场占有率的预测和销售期望利润的预测以及其他商业领域的预测。

三、时间序列模型

1.定义

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。

2.基本特点

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。

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，分析时间序列的方法构成数据分析的一个重要领域，即时间序列分析。

3.常用模型

（1）移动平均法

（2）指数平滑法

（3）差分指数平滑法

（4）平稳时间序列模型

（5）自回归AR

（6）移动平均MA

（7）ARMA模型

4.适用场景

（1）国民经济市场潜量预测

（2）气象预报

（3）水文预报

（4）地震前兆预报

（5）农作物病虫灾害预报

（6）环境污染控制

（7）生态平衡

（8）天文学和海洋学

5.分类

（1）变动形态分类：

长期趋势变动、季节变动、循环变动、不规则变动

（2）方法分类：

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）

移动平均数预测（一次移动平均法、二次移动平均法）

指数平滑法预测（一次、二次、三次指数平滑法）

趋势法预测（分割平均法、最小二乘法、三点法）

季节变动法预测（简单平均法，季节比例法）

四、神经网络预测

1.使用情况：

适用范围很广，但需要有大样本、高算力。