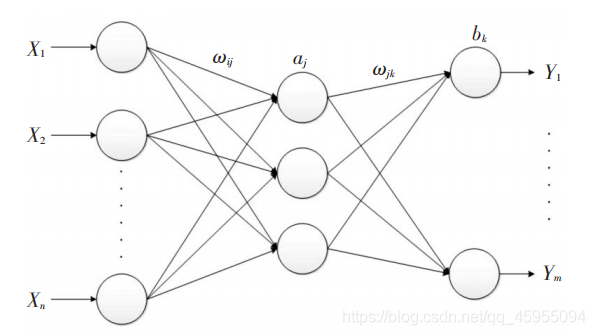
# 基于遗传算法的BP神经网络优化算法

模型的建立

1. BP神经网络预测原理简介

BP 神经网络是一种多层前馈神经网络，常用的为输入层-单隐含层-输出层的三层结构，如下图所示。



BP神经网络训练的主要思想：

1 输入的信号特征数据先映射到隐含层（激活函数实现），再映射到输出层（默认采用线性传递函数），得到期望输出值。

2 将期望输出值和实际测量值做比较，计算误差函数J，再将误差反向传播，通过梯度下降等算法来调节BP网络的权值和阈值。

3 、重复该过程，直到满足设定的目标误差或者最大迭代次数等终止准则，停止训练。

BP神经网络实现了：a). 根据训练集数据，训练得到一个模型，b). 对模型的可靠性与准确性进行测试集（不同于训练样本数据）预测，和实际值对比，检验预测的精度。c). 只给输入，得到预测值（可理解为测试集的数据丢了实测值，本质一样，给输入到BP中，得到输出）

1. 基于遗传算法对神经网络进行优化

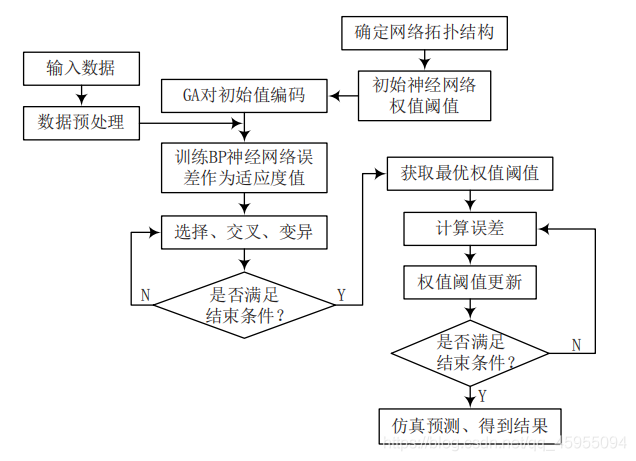
由于其简单的结构，可调整的参数多，训练算法也多，而且可操作性好，BP神经网络获得了非常广泛的应用，但是也存在着一些缺陷，例如学习收敛速度太慢、不能保证收敛到全局最小点、网络结构不易确定。另外，网络结构、初始连接权值和阈值的选择对网络训练的影响很大，但是又无法准确获得，针对这些特点可以采用遗传算法对神经网络进行优化。

在传统BP神经网络训练过程中，通过前向传播数据与误差反向传递，使用算法来更新权重阈值。在该过程中，第一次前向传播过程的权重和阈值该如何确定，即如何初始化权重和阈值。深度学习的方法是采用随机化方法得到初始的权值与阈值参数。选定了初始参数后，梯度下降算法将初始参数值作为起点，进行参数优化与更新。

然后对于优化算法而言，确定性算法，使用数学方法求最优问题，找到的结果与求导的初始点有关，一般为确定值。而对于启发式算法，则是利用某种启发式原则，其主要思想为迭代逼近最优，优化的结果为满足工程精度要求的可变值（无限接近理论最优值）。

对于梯度下降算法而言，其作为一个确定性算法，其收敛性得到证明，但收敛值并非一定是全局最优，与初始的参数值（梯度下降算法的起点）有关。由于随机初始的参数未必是最优的起点（指既训练准确，又预测可靠），因此训练的模型可靠性和稳定性受到了初始随机参数的很大影响。因此可以通过引入启发式算法遗传算法来获得更好的全局搜索能力。

主要思想为将参数作为问题的决策变量，模型的精度作为问题的目标函数。遗传算法GA优化BP神经网络的算法流程图如下：



# 多元线性回归模型

由于客观事物内部规律的复杂性及人们认识程度的限制，无法分析实际对象内在的因果关系，建立合乎机理规律的数学模型。所以在遇到有些无法用机理分析建立数学模型的时候，通常采取搜集大量数据的办法，基于对数据的统计分析去建立模型，其中用途最为广泛的一类随即模型就是统计回归模型。

回归模型确定的变量之间是相关关系，在大量的观察下，会表现出一定的规律性，可以借助函数关系式来表达，这种函数就称为回归函数或回归方程。

回归模型解题步骤：

1. 根据实验数据画出散点图
2. 确定经验公式的函数类型
3. 通过最小二乘法得到正规的方程组
4. 求解方程组，得到回归方程的表达式

建立多元正态线性回归模型本质上是利用不同的方程组确定不同的线性参数

# 随机森林

随机森林的算法属于是组合算法（集成学习）中作为装袋法的突出算法。装袋法的本质是在将多个装入同一个袋子后，让这个袋子作为一个新的模型来实现预测需求。把多个模型组合起来形成一个新的大模型，这个大模型最终给出的预测结果是由这多个小模型综合决定的，决定方式为少数服从多数。

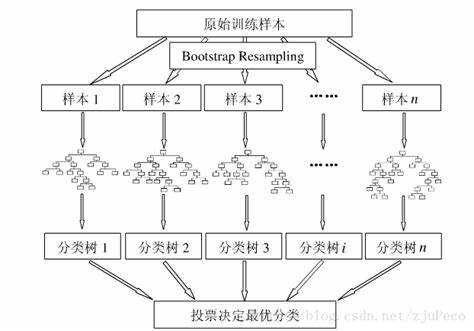
那么给出装袋法的主体算法过程

1. 在样本集上重采样（有重复的）选出n个样本
2. 在所有属性上，对这n个样本建立分类器（ID3、C4.5、CART、SVM、Logistic回归等）
3. 重复以上两步m次，即获得了m个分类器。
4. 将数据放在m个分离器上，最后根据这m个分类器的投票结果，决定数据属于哪一类

同时在此基础上给出随机森林的实现步骤：

1. 使用装袋法在行列上进行随机抽样
2. 有很多决策树分类器组合而成（因而称为“森林”）
3. 单个的决策树分类器使用随机方法构成。
   1. 首先，学习集是从原训练集中通过有放回抽样得到的自助样本。
   2. 其次，参与构建该决策树的变量也是随机取出，参与变量数通常大多小于可用变量数（即第一点上的在列上而随机抽样）
4. 单个决策树在产生学习集和确定参与变量后，使用CART算法计算，不会剪枝（因为无需考虑过渡拟合的问题）
5. 最后分类结果取决于各个决策树分类器简单多数选举

随机森林模型具有着多种优点：其准确性明显高于组合中任何单个分类器，准确率达到极大提高；其对于错误和离群点，也可被视为噪声，不会表现很差，且具有鲁棒性；决策树容易过度拟合的问题会随着森林的规模而削弱；大数据情况下速度快（分布式），性能号



# Xgboost模型

Xgboost（Extreme Gradient Boost）模型，是一种特殊的梯度提升决策树（GBDT，Gradient Boosting Decision Tree），只不过是力求将速度和效率发挥到了极致，故叫X（Extreme）gradientboost。Xgboost其本质上还是基于树结构并结合集成学习的一种方法，其基础树结构为分类回归树（CART，Classification and Regression Tree）。类似于局部加权线性回归算法，基于树的回归算法也是一类局部的回归算法，通过将数据集切分成多份，在每一份数据上单独建模。

CART（分类回归树）算法核心内容包含以下三个方面：

1. 二分(Binary Split)：在每次判断过程中，都是对观察变量进行二分。CART算法采用一种二分递归分割的技术，算法总是将当前样本集分割为两个子样本集，使得生成的决策树的每个非叶结点都只有两个分枝。因此CART算法生成的决策树是结构简洁的二叉树;
2. 单变量分割(Split Based on One Variable)：每次最优划分都是针对单个变量;
3. 剪枝策略：CART算法的关键点，也是所有基于树（Tree-Based）算法的关键步骤

决策树的生成就是递归地构建二叉决策树的过程，核心思想为在训练数据集所在的输入空间中，递归地将每个区域划分为两个子区域并决定每个子区域上输出值。划分子区域的标准取决于树的种类，对回归树用平方误差最小化准则，对分类树用基尼指数最小化准则。回归树的生成具体步骤如下：

1. 选择最优切分变量j与切分点s，求解

*min*[*minxi*​∈*R*1​(*j*,*s*)∑​(*yi*​−*c*1​)2+*minxj*​∈*R*2​(*j*,*s*)∑​(*yj*​−*c*2​)2]

遍历变量j，对固定的切分变量j扫描切分点s选择使上式最小值的对(j;s)。其中Rm是被划分的输入空间，cm是空间Rm对应的输出值

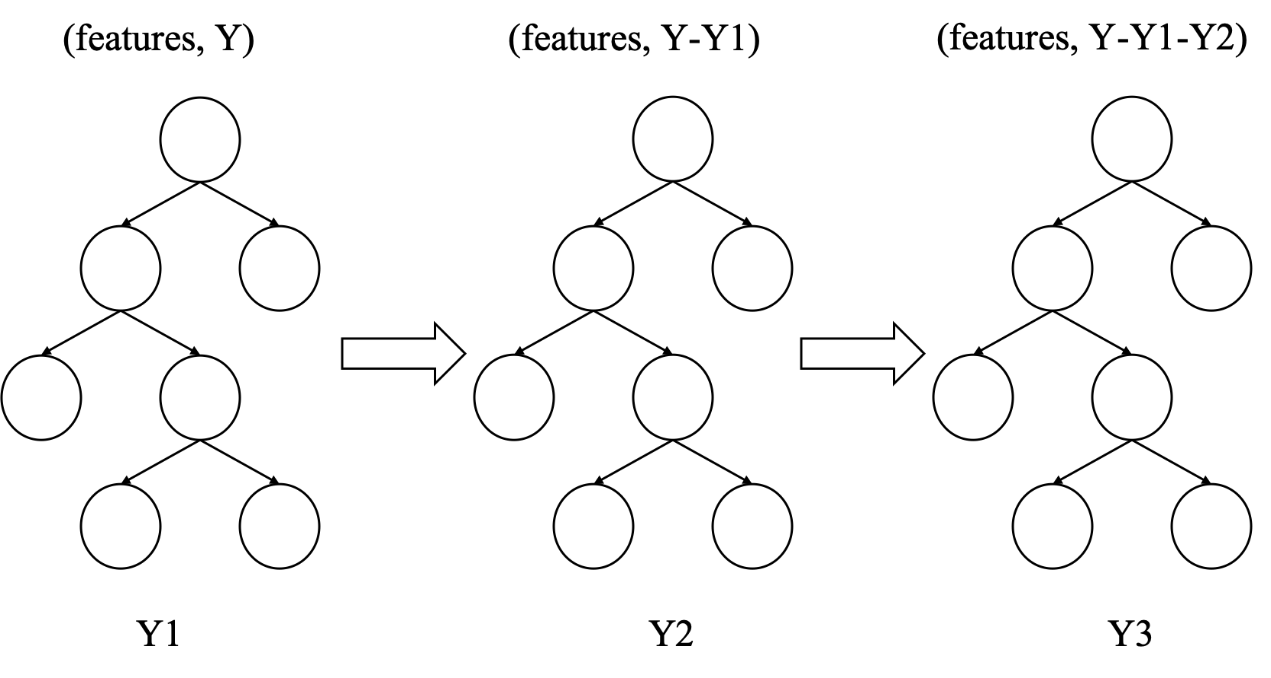
1. 用选定的对 (j;s)划分区域并决定相应的输出值

*R*1​(*j*,*s*)={*x*∣*x*(*j*)≤*s*},*R*2​(*j*,*s*)={*x*∣*x*(*j*)>*s*}

*c*^*m*​=*Nm*​1​*xi*​∈*Rm*​(*j*,*s*)∑​*yi*​,*x*∈*Rm*​,*m*=1,2

1. 继续对两个子区域递归的调用上述步骤，最终将输入空间划分为M个区域*R*1​,*R*2​,…,*Rm*​，生成决策树
2. 当输入空间划分确定时，可以用平方误差来衡量回归树对于训练数据的拟合程度，用平方误差最小的准则不断地递归划分子树，直到平方误差满足需求

单棵分类回归树精度有限，应用场景受限，故Xgboost在CART的基础上引入了集成学习（boosting方法），并采用并行计算等方式极大的加速了模型计算速度。Boosting的核心思想就是所有弱分类器的结果相加等于预测值，然后下一个弱分类器去拟合误差函数对预测值的梯度/残差(这个梯度/残差就是预测值与真实值之间的误差)，从而不断地减小残差，直到满足系统的误差要求（如图所示）



# 决策树

决策树是一种从无次序、无规则的样本数据集中推理出决策树表示形式的分类规则方法。决策树学习的算法通常是一个递归地选择最优特征，并根据该特征对训练数据进行分割，使得各个子数据集有一个最好的分类的过程。

在决策树算法中有一个非常重要的概念：信息熵

信息熵是信息论中用于度量信息量的一个概念。一个系统越是有序，信息熵就越低；反之，一个系统越是混乱，信息熵就越高。所以，信息熵也可以说是系统有序化程度的一个度量。

信息熵的计算公式

在这里插入图片描述

其中xi 是变量。p(xi)是变量xi发生的概率

目标是创建一个模型，通过从数据特性中推导出简单的决策规则来预测目标变量的值。

基于因子分析法与熵权法的模糊综合评价法

1. 因子分析法

因子分析是指研究从变量群中提取共性因子的统计技术。因子分析可在许多变量中找出隐藏的具有代表性的因子。将相同本质的变量归入一个因子，可减少变量的数目，还可检验变量间关系的假设。

简而言之，因子分析法，认为很多变量之间具有相关性，而这种相关性本质上是指多个变量可能存在着一个共同的影响因子，这一个影响因子就可以通过多个函数得到多个变量的近似。那么我们就可以用这一个隐变量去表示多个变量的信息。

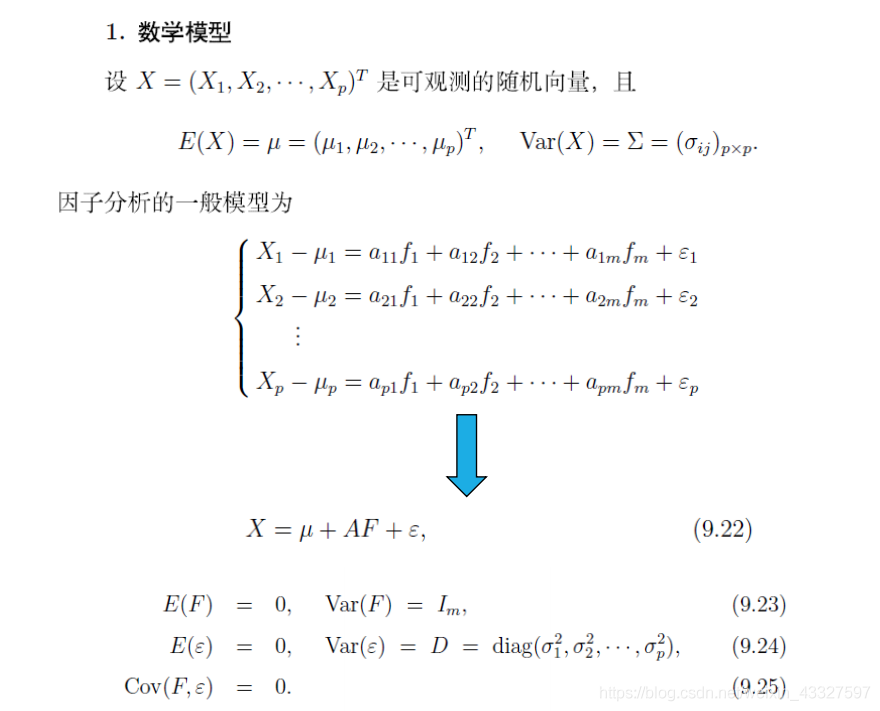
过程步骤

1. 确定原有若干变量是否适合于因子分析；因子分析的基本逻辑是从原始变量中构造出少数几个具有代表意义的因子变量，这就要求原有变量之间要具有比较强的相关性，否则，因子分析将无法提取变量间的“共性特征”
2. 构造因子变量；因子分析中有多种确定因子变量的方法，如基于主成分模型的主成分分析法和基于因子分析模型的主轴因子法、极大似然法、最小二乘法等。
3. 利用旋转使得因子变量更具有可解释性 ；在实际分析工作中，主要是因子分析得到因子和原变量的关系，从而对新的因子能够进行命名和解释，否则其不具有可解释性的前提下对比PCA就没有明显的可解释价值。
4. 计算因子变量的得分。子变量确定以后，对每一样本数据，希望得到它们在不同因子上的具体数据值，这些数值就是因子得分，它和原变量的得分相对应。

具体步骤：

1. 相关性检验，一般采用KMO检验法和Bartlett球形检验法两种方法来对原始变量进行相关性检验；
2. 相关性检验，一般采用KMO检验法和Bartlett球形检验法两种方法来对原始变量进行相关性检验；
3. 计算样本的相关矩阵R；
4. 求相关矩阵R的特征根和特征向量；
5. 根据系统要求的累积贡献率确定公共因子的个数；
6. 计算因子载荷矩阵A；
7. 对载荷矩阵进行旋转，以求能更好地解释公共因子；
8. 确定因子模型；
9. 根据上述计算结果，求因子得分，对系统进行分析

数学模型的公式化表示



1. 基于熵权法的模糊评价法

利用模糊数学理论,结合预测评价指标,针对现有方法的优点和存在的问题,利用评价模型的优劣程度由轻到重逐渐变化的模糊特性,可以获得更科学和更合理的评价结果. 在水质模糊综合评价中,需确定影响模型优劣的各主要因素,确定评价因子集、评价集、隶属函数,然后通过计算各因素的权重和隶属度,得到综合隶属度,确定预测级别.

在模糊综合评价中,权重的设计是一项重要的内容,对评价的结果有重要影响. 现有模糊综合评价方法一般是通过计算超标比来确定各因子的权重,即根据某待评价对象的各评价指标的监测值相对于标准均值的超标程度经过归一化计算所获得的结果作为因子权重[4 ] . 此方法存在以下不足之处:当存在多个评价对象时,每个评价对象都要分别计算一次在该对象下每个评价指标的权重值,工作量过大;计算得到的权重值仅考虑了个体因子的特征,而对多个评价对象的相互联系却无法描述. 因此,现行的模糊综合评价方法在预测模型评价中受到限制.

为解决以上问题,可将熵的概念应用到水质评价中. 熵本是热力学中的概念, 后由申农(C. E.Shannon) 引入信息论,现已在工程技术、社会经济等领域得到广泛应用[5～14 ]

人们在评价决策中所获信息的多少,是评价精度和可靠性大小的决定因素之一. 在信息论中,熵是系统无序程度的度量,它还可以度量数据所提供的有效信息量[15] . 因此,可以用熵来确定权重.当评价对象在某项指标上的值相差较大时,熵值较小,说明该指标提供的有效信息量较大,该指标的权重也应较大;反之,若某项指标的值相差越小,熵值较大,说明该指标提供的信息量较小,该指标的权重也应较小. 当各被评价对象在某项指标上的值完全相同时,熵值达到最大,这意味着该指标未向决策提供任何有用的信息,可以考虑从评价指标体系中去除[16] .所以“, 熵权”理论是一种客观赋权方法.在水质模糊评价中,通过对“熵”的计算确定权重,就是根据各项监测指标值的差异程度,确定各指标的权重.

模型整体过程

1. 将原始矩阵进行标准化，消除评价指标的量纲所导致的差异性；
2. 计算不同指标的信息熵熵值
3. 获取指标权值
4. 利用熵权法对模糊评价法进行优化，定权
5. 计算评价度S
6. 最后得到各个模型的综合评价结果

需要说明的是,熵权并不是表示决策评估问题中某指标实际意义上的重要性系数,而是在给定评价对象集后各种评价指标值确定的情况下,各指标在竞争意义上的相对激烈程度. 从信息角度考虑,它代表该指标在该问题中,提供有效信息量的多寡程度[16] .

优点：

与传统的模糊综合评价方法相比,通过熵权法给各评价因子赋权，只需计算1 次就能得到适用于所有点的权重,大大减少了评价工作所需的工作量.

使用熵权法赋权,可将同一监测指标的多个样本点结合确定权重,考虑了多个样本间的联系,可削弱异常值的影响,使评价结果更准确、合理.

实例计算结果表明,熵权法是一种比较有效的赋权方法,在预测模糊综合评价中有重要应用价值.

[ 5 ] 　Tian Q H , Du Y X. Study of Performance Evaluation for Mechanical Products Based on Entropy Fuzzy Comprehensive Review[J ] . China Manufacturing Information , 2004 , 33 (3) : 97 —99 (in Chinese)

[15 ] 　Meng Q S. Information Theory [ M] . Xi′An : Xi′An Jiaotong University Press , 1989 :19 —36 (in Chinese)

[16 ] 　Qiu W H. Management Decision and Applied Entropy[M] . Beijing :China Machine Press , 2002 : 193 —196 (in Chinese)