

# 如何定量地比较二个分类器的优劣-大学物理实验中 怎样比较两组测量数据的优劣 求解-股识吧

## 一、线性分类器在联合算法当中性能和不稳定性如何描述？

展开全部这个问题不算很前沿，但是相当深奥，在目前国际上存在一定争议，接受比较广泛的定义是一个基于实验的研究文献，我的答案源于翻译的文献，您在SCI当中搜索线性分类器联合算法，影响因子最大的就是它的原文。

研究线性分类器联合技术的有用性与它们的不稳定性和训练样本之间的关系，让我们考虑一些线性分类器。

最流行也是最普遍的是费希尔线性判别（FLD）和最近均值分类器（NMC），也被叫做欧氏距离分类器。

标准FLD定义为： $gF(x) = x' wF + w0F = [X - 1/2(X(1) + X(2))] ' S^{-1}(X(1) - X(2))$ 其中 $(wF, w0F)$ 是线性判别函数的系数， $S$ 是 $p$ 维协方差矩阵的标准最大似然估计， $x$ 是待分类的 $p$ 维向量， $X(i)$ 是类别  $i$ 的样本均值向量。

注意到等式1是线性系数 $(w, w0)$ 的均方误差解： $gF(x) = w \cdot x + w0 = L$ 其中 $x \in X$ ， $L$ 是理想输出， $1$ 对应类别  $1$ ， $-1$ 对应类别  $2$ 。

当特征数 $p$ 超过样本 $n$ 的时候，协方差矩阵的样本估计 $S$ 是不可逆的奇异阵。

当特征数 $p$ 增加到 $n$ 的时候，误差的期望显著增加。

NMC定义为： $gPFLD(x) = (w, w0) \cdot (x, 1) = (x, 1)(X, I) - 1L$ 它是向量 $x$ 和样本均值 $X(i)$ 之间的最小距离。

NMC生成样本均值的中垂线，因而生成一种类似于具有等方差高斯球分布类别的理想线性分类器。

这种分类器的优点在于它对于样本数不是很敏感。

然而，NMC不考虑方差和协方差。

FLD的修正型让我们避免了病态协方差矩阵的求逆，被称为伪费希尔线性判别（PFLD）。

在PFLD中，获得了一种直接解法（通过增广向量）： $gPFLD(x) = (w, w0) \cdot (x, 1) = (x, 1)(X, I) - 1L$ 其中 $(x, 1)$ 是增广向量， $(X, I)$ 是增广矩阵， $(X, I)^{-1}$ 是莫奈-罗斯伪逆，给出了最低标准解。

在求逆之前数据被转换为零均值，这种方法与奇异值分解密切相关。

作为一个训练样本的函数来研究PFLD的性能不在此讨论，对于每类的一个样本，这种方法等价于最近均值和最近邻准则。

对于值 $n >$ ；

$p$ 的PFLD，所有给定样本的总距离的最大化，等价于FLD。

然而，对于值 $n = p$ 的情况，伪费希尔决策找到了一个涵盖所有数据的线性空间。

在两者之间，PFLD的普遍误差最小，但是随即在 $n = p$ 处变为最大。

这可以理解为在 $n=p$ 之前，PFLD可以成功的找到所有训练样本的等距离超球面。得到了精确解，但是涵盖了所有的数据误差。

在Raudys和Duin的文献当中，从PFLD的普遍误差到处了一个渐进表达式，从理论上解释了PFLD的走势。

PFLD的优点在于它简约明了，简单的分类决策，以及考虑到不同的方差。

然而，当建立在边界样本的时候，它非常弱并且对噪声很敏感。

在这种情况下，不建议作为单一分类器来使用它。

不过，由于PFLD普遍误差的峰值现象，这种分类器非常的适合研究维度（样本大小）影响。

在这里，将被用作联合技术的基础分类器。

为了更好地理解联合技术何时有益，考虑基础分类器的不稳定性。

分类器的不稳定性，通过计算原始训练集的引导副本所引起的测试集分类变动来作为衡量依据。

在训练集中重复此过程多次（我们做到了25次），对结果求平均，可以得到一个该训练集分类不稳定的估计。

以这种方式定义的线性分类器平均不稳定性（50个独立训练集）。

人们可以看到，当训练样本大小增加的时候分类的不稳定性降低。

分类的不稳定性和性能之间的一个关系：更稳定的分类比不太稳定的执行好。

然而，在这个例子中，NMC的分类性能并不取决于训练样本的大小。

与其他分类相比，在大样本的情况下，随着稳定性增加，它仍然是性能很差的分类器。

## 二、在仪器分析中用什么指标来衡量一个检测器的优劣

事实上这个题目如果不指明什么仪器和检测器的话，这道题很难答。

并且不知道所谓的检测器是指仪器的检测器部分还是整个仪器为检测器。

只能大而概之地回答。

- 1、线性范围，即一个检测器斜率保持不变的浓度或者质量范围。
- 2、检测限，检测器所能检测到的最低浓度或者质量。
- 3、测量限，检测器所能准确测量的最低浓度或质量。
- 4、灵敏度，在GC中指响应信号对进样量的变化率，一般指的是直线的斜率。
- 5、选择性，在GC中指被测组分的响应值与干扰组分的响应值之比，一般指检测器抗干扰能力。
- 6、精密度，这个包括方法的精密度以及仪器的精密度，通常无法区分。
- 7、稳定性，在电位法中通常用稳定性用恒定条件下，电极系统的电位漂移来表示。

。

8、重现性，在电位法中指电极往复三次测定浓度为 $10^{-3}$  mol/L和 $10^{-2}$  mol/L的溶液时所得的电位平均偏差。

### 三、求对比这两个滤波器的优劣（图）

### 四、能给我描述一下制度管理和流程管理的关系及优缺点么

制度的产生是人性善恶的不同选择，之所以有制度，是因为对人性“恶”的假设。当企业达到一定规模时，管理者无法亲自照看到企业的一个角落，于是就有了制度。

而流程管理的思想则强调激发人性“善”的一面，从而实现持续的改进。

两者的对立统一主要表现在：制度往往强调“做了某一事情的后果的处理方法”，其内容是一种奖惩的规定，彼此之间相对独立。

而流程则强调如何将输入有效地转化为输出，强调“如何去把一件事情做得更好”。

流程就像是河流，流程管理就像是河道梳理，制度就是巩固河道的堤坝，那如果是治理象河流一样的业务流程，那应该用流程管理先梳理河道，河道梳理好了以后再管理制度这样的堤坝巩固梳理成果，如果只是加固堤坝，河流总有一天会泛滥成灾的。

不同的企业适用于不同的管理模式，而采用“制度导向”的管理还是“流程导向”的管理，取决于管理者所基于的不同假设“人性本善”和“人性本恶”。

“制度导向”是采取“以堵治水”的办法，基于的管理假设是“人性本恶”。

简单的说既是“假如你不做，按制度进行处罚；

你要胆敢再犯，加倍处罚；

看你还敢吗？”。

而“流程导向”更多的是一种“以导治水”，其基于的管理假设是“人性本善”，特点是以完成工作步骤、顺序作为核心，结合组织结构、人员素质及其他资源，站在公司的角度，来设定流程；

它是企业管理原则的很好载体；

它提倡以“对自己职责的本分”、“对上下游的积极信任”的态度来有效运作流程。

制度与流程的关系还表现在局部观念与全局观念的不同，从“流程”的定义中可以了解到“流程导向”是为实现某项功能的一个系统，系统可大可小，整个企业是一

个系统，根据不同的分类原则内部又可以分为若干个独立系统，各系统之间都会通过各流程系统之间的接口建立起紧密联系，最终织成一个涵盖全局的网络系统。

而“制度导向”更多的是针对局部出现执行力问题而采取的奖惩措施，包括对执行人主观态度以及客观过失造成公司损失的处理。

流程是制度的灵魂：如果制度不能反应流程，就像失去了灵魂，它的执行一定会出现问题。

所以制度无法执行时，往往是它所包含的流程有问题。

制度是流程得以执行的保证：制度是因流程而存在的，通过适当制度的执行推动流程的执行。

流程是建立在对功能团队信任的基础上而设计的，对于因个体原因而影响流程功能实现的现象，只有通过制度进行约束，才能得以制止，进而建立流程的威信。

## 五、4.如何将两类判别法(如支持向量机)应用到多类问题?

这个一般有两种方法：给你个我看的别个总结的：我们就会想到所谓“一类对其余”的方法，就是每次仍然解一个两类分类的问题。

比如我们有5个类别，第一次就把类别1的样本定为正样本，其余2, 3, 4, 5的样本合起来定为负样本，这样得到一个两类分类器，它能够指出一篇文章是还是不是第1类的；

第二次我们把类别2的样本定为正样本，把1, 3, 4, 5的样本合起来定为负样本，得到一个分类器，如此下去，我们可以得到5个这样的两类分类器（总是和类别的数目一致）。

到了有文章需要分类的时候，我们就拿着这篇文章挨个分类器的问：是属于你的么？是属于你的么？哪个分类器点头说是了，文章的类别就确定了。

这种方法的好处是每个优化问题的规模比较小，而且分类的时候速度很快（只需要调用5个分类器就知道了结果）。

但有时也会出现两种很尴尬的情况，例如拿一篇文章问了一圈，每一个分类器都说它是属于它那一类的，或者每一个分类器都说它不是它那一类的，前者叫分类重叠现象，后者叫不可分类现象。

分类重叠倒还好办，随便选一个结果都不至于太离谱，或者看看这篇文章到各个超平面的距离，哪个远就判给哪个。

不可分类现象就着实难办了，只能把它分给第6个类别了.....更要命的是，本来各个类别的样本数目是差不多的，但“其余”的那一类样本数总是要数倍于正类（因为它是除正类以外其他类别的样本之和嘛），这就人为的造成了上一节所说的“数据集偏斜”问题。

因此我们还得再退一步，还是解两类分类问题，还是每次选一个类的样本作正类样

本，而负类样本则变成只选一个类（称为“一对一单挑”的方法，哦，不对，没有单挑，就是“一对一”的方法，呵呵），这就避免了偏斜。

因此过程就是算出这样一些分类器，第一个只回答“是第1类还是第2类”，第二个只回答“是第1类还是第3类”，第三个只回答“是第1类还是第4类”，如此下去，你也可以马上得出，这样的分类器应该有 $5 \times$

$4/2=10$ 个（通式是，如果有 $k$ 个类别，则总的两类分类器数目为 $k(k-1)/2$ ）。

虽然分类器的数目多了，但是在训练阶段（也就是算出这些分类器的分类平面时）所用的总时间却比“一类对其余”方法少很多，在真正用来分类的时候，把一篇文章扔给所有分类器，第一个分类器会投票说它是“1”或者“2”，第二个会说它是“1”或者“3”，让每一个都投上自己的一票，最后统计票数，如果类别“1”得票最多，就判这篇文章属于第1类。

这种方法显然也会有分类重叠的现象，但不会有不可分类现象，因为总不可能所有类别的票数都是0。

看起来够好么？其实不然，想想分类一篇文章，我们调用了多少个分类器？10个，这还是类别数为5的时候，类别数如果是1000，要调用的分类器数目会上升至约500,000个（类别数的平方量级）。

这如何是好？

## 六、求对比这两个滤波器的优劣（图）

两者性能没有区别，下面一个可以比上面一个电路节省一个电阻。

工程上从不自己计算滤波器参数，直接查手册一切OK。

推荐电子书《有源滤波器的快速实用设计》、《有源滤波器精确设计手册》，下载后自己查找。

## 七、大学物理实验中 怎样比较两组测量数据的优劣 求解

大学物理实验中 怎样比较两组测量数据的优劣从四个方面来回答，1．设计类型是完全随机设计两组数据比较，不知道数据是否是连续性变量。

2．比较方法：如果数据是连续性数据，且两组数据分别服从正态分布&方差不齐（方差齐性检验），则可以采用t检验，如果不服从以上条件可以采用秩和检验。

3．想知道两组数据是否有明显差异？不知道这个明显差异是什么意思？是问差别有无统计学意义（即差别的概率有多大）还是两总体均数差值在哪个范围波动？如果

是前者则可以用第2步可以得到P值，如果是后者，则是用均数差值的置信区间来完成的。

当然两者的结果在SPSS中均可以得到。

4. 对以上结果SPSS的实现是：（1）t检验，analyse compare means independent-samples T Test （2）秩和检验，analyse noparametric Test 2 independent samples

## 参考文档

[下载：如何定量地比较二个分类器的优劣.pdf](#)

[《支付宝买股票模拟买入是什么意思》](#)

[《什么炒股软件速度最快》](#)

[《matlab怎么接入股票数据》](#)

[《股票突然st怎么办》](#)

[下载：如何定量地比较二个分类器的优劣.doc](#)

[更多关于《如何定量地比较二个分类器的优劣》的文档...](#)

声明：

本文来自网络，不代表

【股识吧】立场，转载请注明出处：

<https://www.gupiaozhishiba.com/author/33333336.html>