这里换成你的论文的标题

摘 要

开头段：需要充分概括论文内容，一般两到三句话即可，长度控制在三至五行。

问题一中，解决了什么问题；应用了什么方法；得到了什么结果。

问题二中，解决了什么问题；应用了什么方法；得到了什么结果。

问题三中，解决了什么问题；应用了什么方法；得到了什么结果。

结尾段：可以总结下全文，也可以介绍下你的论文的亮点，也可以对类似的问题进行适当的推广。

关键词：关键词1 关键词2 关键词3 关键词4

|  |
| --- |
| **注意事项：**  本Word模版的版本编号是0.1版，是以国赛的标准创建的模版，也适用于国内绝大多数的数学建模比赛。模版的使用方法可以查看下面这个视频：  数学建模清风——论文排版教程  <https://www.bilibili.com/video/BV1Ci4y1c7Ld>  未来当发现模版中的问题或者比赛要求有更新时，我会发布更新后的新版本。大家可以在公众号《数学建模学习交流》后台发送“国赛论文模版”获取最新的模版的信息。只要本文档的版本号和公众号后台最新的版本号一致，则说明你下载的是最新版本。  关于具体怎么写好数学建模论文的每一部分，可以看下面这个视频：  <https://www.bilibili.com/video/BV1Na411w7c2/>  **红色字体的文字是上面这个视频中的笔记，在实际论文中不要出现。**  首页三要素: 论文标题 + 摘要 + 关键词  （1）标题：   * 基于所使用的主要模型或者方法作为标题（推荐） * 直接使用赛题所给的题目或者要研究的问题作为标题   （2）摘要：  摘要是数模论文写作中最重要的一部分，因为评阅老师的时间有限，拿到一篇论文后不会完整的从头读到尾，所以评阅老师往往会重点阅读摘要部分，并结合官方的评阅要点来对你的论文进行初步评定。因此，大家一定要好好打磨论文的摘要，摘要一般是其他部分都完成后再来书写，写完后需要反复阅读反复修改。  （3）关键词：  关键词一般放4-6个，可以放论文中使用的主要模型，也可以放论文里面出现次数较多，能体现论文的主要内容的词。 |

# 问题重述

数学建模比赛论文是要我们解决一道给定的问题，所以正文部分一般应从问题重述开始，一般确定选题后就可以开始写这一部分了。

这部分的内容是将原问题进行整理，将问题背景和题目分开陈述即可，所以基本没啥难度。

本部分的目的是要吸引读者读下去，所以文字不可冗长，内容选择不要过于分散、琐碎，措辞要精练。

注意：在写这部分的内容时，绝对不可照抄原题！（论文会查重）

应为：在仔细理解了问题的基础上，用自己的语言重新将问题描述一遍。语言需要简明扼要，没有必要像原题一样面面俱到。

# 问题分析

## 问题一的分析

从实际问题到模型建立是一种从具体到抽象的思维过程，问题分析这一部分就是沟通这一过程的桥梁，因为它反映了建模者对于问题的认识程度如何，也体现了解决问题的雏形，起着承上启下的作用，也很能反应出建模者的综合水平。

这部分的内容应包括：题目中包含的信息和条件，利用信息和条件对题目做整体分析，确定用什么方法建立模型，一般是每个问题单独分析一小节，分析过程要简明扼要， 不需要放结论。

建议在文字说明的同时用图形或图表（例如流程图）列出思维过程，这会使你的思维显得很清晰，让人觉得一目了然。

（注意：问题分析这一部分放置的位置比较灵活，可以放在问题重述后面作为单独的一节(见到的频率最高)，也可以放在模型假设和符号说明后面作为单独的一节，还可以针对每个问题将其写在模型建立中。具体可以看视频讲解）

## 问题二的分析

## 问题三的分析

# 模型假设

视频中介绍了6类常见的模型假设：

1. 题目明确给出的假设条件
2. 排除生活中的小概率事件(例如黑天鹅事件、非正常情况)
3. 仅考虑问题中的核心因素，不考虑次要因素的影响
4. 使用的模型中要求的假设
5. 对模型中的参数形式(或者分布)进行假设
6. 和题目联系很紧密的一些假设，主要是为了简化模型

# 符号说明

|  |  |  |
| --- | --- | --- |
| **符号** | **说明** | **单位** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

本部分是对模型中使用的重要变量进行说明，一般排版时要放到一张表格中。

注意：第一：不需要把所有变量都放到这个表里面，模型中用到的临时变量可以不放。第二：下文中首次出现这些变量时也要进行解释，不然会降低文章的可读性。

# 问题一模型的建立与求解

## 卡方分析

卡方分析是用来研究两个定类变量间是否独立即是否存在某种关联性的最常用的方法[[1]](#footnote-1)。卡方值的计算公式如下：

|  |  |
| --- | --- |
|  | () |

其中O代表观测值，E代表期望值。

1. 第一问

我们利用SPSSAU分别对玻璃文物的表面风化与玻璃类型、纹饰和颜色进行卡方分析，即判断其是否存在一定关联度，下表给出了卡方分析的结果。

表 卡方分析结果表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 题目 | 名称 | 表面风化(%) | | 总计 |  |  |
| 无风化 | 风化 |
| 类型 | 铅钡 | 12(50.00) | 28(82.35) | 40(68.97) | 6.880 | 0.009\*\* |
| 高钾 | 12(50.00) | 6(17.65) | 18(31.03) |
| 总计 | | 24 | 34 | 58 |  |  |
| 颜色 | 无 | 0(0.00) | 4(11.76) | 4(6.90) | 9.432 | 0.307 |
| 浅绿 | 2(8.33) | 1(2.94) | 3(5.17) |
| 浅蓝 | 8(33.33) | 12(35.29) | 20(34.48) |
| 深绿 | 3(12.50) | 4(11.76) | 7(12.07) |
| 深蓝 | 2(8.33) | 0(0.00) | 2(3.45) |
| 紫 | 2(8.33) | 2(5.88) | 4(6.90) |
| 绿 | 1(4.17) | 0(0.00) | 1(1.72) |
| 蓝绿 | 6(25.00) | 9(26.47) | 15(25.86) |
| 黑 | 0(0.00) | 2(5.88) | 2(3.45) |
| 总计 | | 24 | 34 | 58 | 4.957 | 0.084 |
| 纹饰 | A | 11(45.83) | 11(32.35) | 22(37.93) |
| B | 0(0.00) | 6(17.65) | 6(10.34) |
| C | 13(54.17) | 17(50.00) | 30(51.72) |
| 总计 | | 24 | 34 | 58 |  |  |

**注：\* p<0.05 \*\* p<0.01**

从上表可以看出：不同表面风化样本对于颜色和纹饰不会表现出显著性(p>0.05)，意味着不同表面风化样本对于颜色和纹饰均表现出一致性，并没有差异性。另外表面风化样本对于类型项呈现出显著性(p<0.05)，意味着不同表面风化样本对于类型项均呈现出差异性。

表面风化对于类型呈现出0.01水平显著性(chi=6.880, p=0.009<0.01)，通过百分比对比差异可知，风化选择铅钡的比例82.35%，明显高于无风化的选择比例50.00%。无风化选择高钾的比例50.00%，明显高于风化的选择比例17.65%。

总结可知：不同表面风化样本对于颜色和纹饰不会表现出显著性差异，另外，表面风化样本对于类型项呈现出显著性差异。

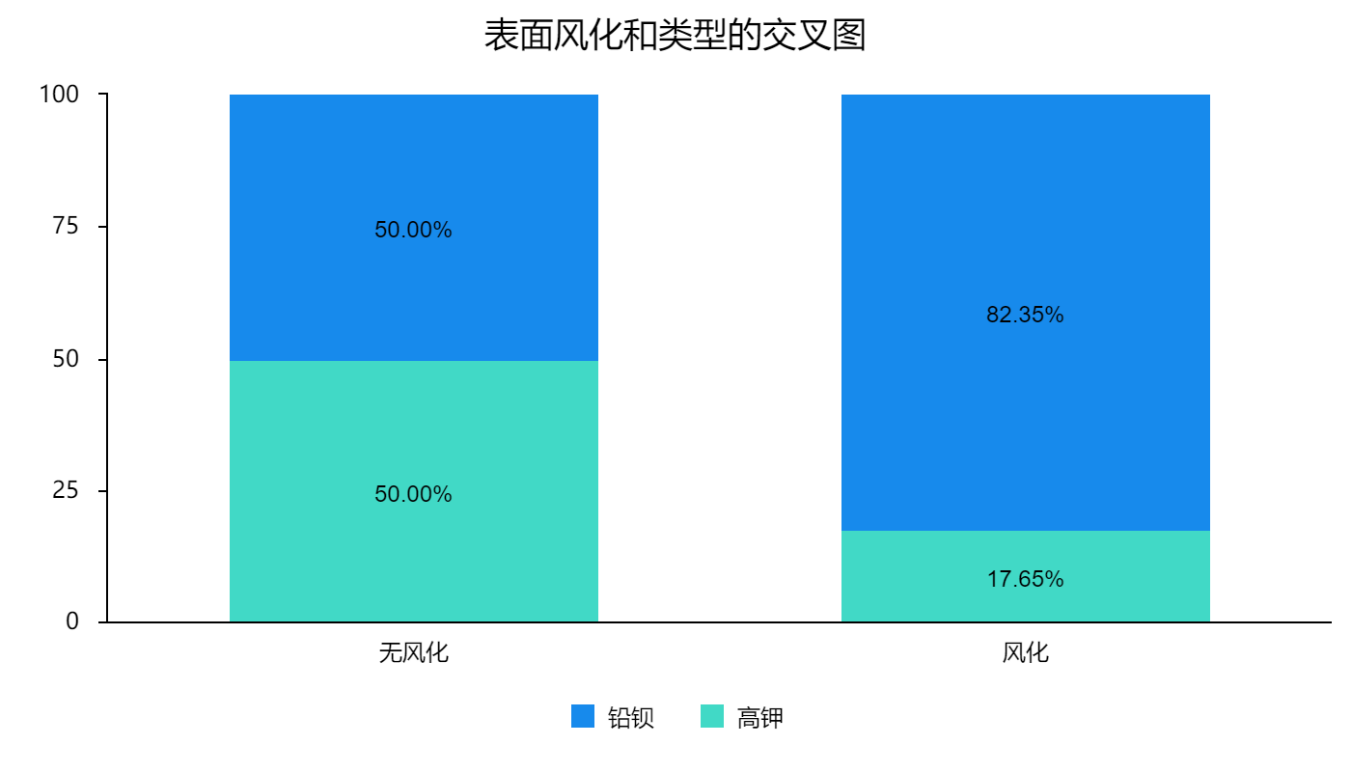


图 表面风化和类型的堆积柱形图

# 问题二模型的建立与求解

## 第二问

# 问题三模型的建立与求解

针对问题三，我们需要利用附件.xlsx中的表单2的有效数据对未知类型玻璃文物进行类型鉴定。对此，问题三可以定义为一个分类问题，可以采用机器学习方法对其进行建模和求解。但我们也注意到，有效数据集大小仅为67，一般的机器学习算法和工具都集中在“大数据”和具有大量数据集的场景中，但是本题可利用的数据集比较有限，一般的算法处理起来比较棘手，容易出现过拟合问题。

为了解决小数据集容易给机器学习算法带来过拟合的问题，我们建立了一个简单的分类模型，并对原特征进行降维处理，使用较少的特征进行分类，以缓解小数据集带来的过拟合问题。

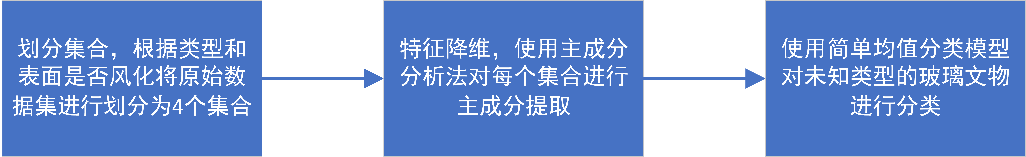


图 问题三求解流程图

## 划分集合

我们考虑到类型和表面是否风化对玻璃文物的化学成分影响较大，故在提取主成分前，我们根据类型和表面是否风化将数据分为了高钾-风化、高钾-未风化、铅钡-风化和铅钡-未风化四类，然后利用PCA分别对这四类数据集合进行特征提取、降维。

## 主成分分析特征降维

特征选择对于建立模型尤为重要。尽管正则化之类的方法有助于减少特征，但是如果特征数远远大于样本数，那么过拟合的问题仍然会持续存在。为了缓解小数据集、多特征项带来的过拟合的问题，我们首先通过主成分分析法对数据特征进行降维，将原本的14个化学成分（特征）提取为少数几个可以反映出原始数据大部分信息（累积方差解释率超过80%）的主成分特征，然后以提取出的主成分特征进行下一步的简单均值分类。下表展示了高钾-风化类的PCA线性组合系数表，其被提取为了2个主成分特征，特征数减少了12个。（由于表格所占篇幅较大，我们会在附录中放入其他3类的PCA线性组合系数表）

表 高钾-风化类的PCA线性组合系数表

|  |  |  |
| --- | --- | --- |
| **名称** | **成分** | |
| **成分1** | **成分2** |
| 二氧化硅(SiO2) | -0.475 | -0.093 |
| 氧化钾(K2O) | -0.386 | 0.284 |
| 氧化钙(CaO) | 0.402 | 0.287 |
| 氧化镁(MgO) | 0.294 | 0.406 |
| 氧化铝(Al2O3) | 0.396 | 0.342 |
| 氧化铁(Fe2O3) | -0.182 | 0.452 |
| 氧化铜(CuO) | 0.208 | -0.493 |
| 五氧化二磷(P2O5) | 0.379 | -0.316 |

## 简单启发式均值分类模型

使用更简单的模型，是因为它们不太容易过拟合，比如正则化线性模型，弹性网络分类器等等。然而线性模型虽然简单，但是在特征维数较多的小数据集上的分类准确率往往不尽人意。

对此，我们一开始想到使用较为简单的K-NN模型来进行分类，其原理就是给定一个已知标签类别的训练数据集，输入没有标签的新数据后，在训练数据集中找到与新数据最邻近的k个实例，如果这k个实例的多数属于某个类别，那么新数据就属于这个类别[[2]](#footnote-2)。

但由于通过类型和表面是否风化对数据进行划分后的每个类别的数据数量不一致，这在一定程度上会影响K-NN分类出来结果的准确性。于是，我们借鉴K-NN分类的思想，提出了利用每个类型的特征（这里的特征指的是PCA提取出的主成分特征）均值进行分类的模型，并在其中加入了一些启发式规则来提升分类的准确率，称之为简单启发式均值分类模型。下图展示了简单启发式均值分类模型的流程图：

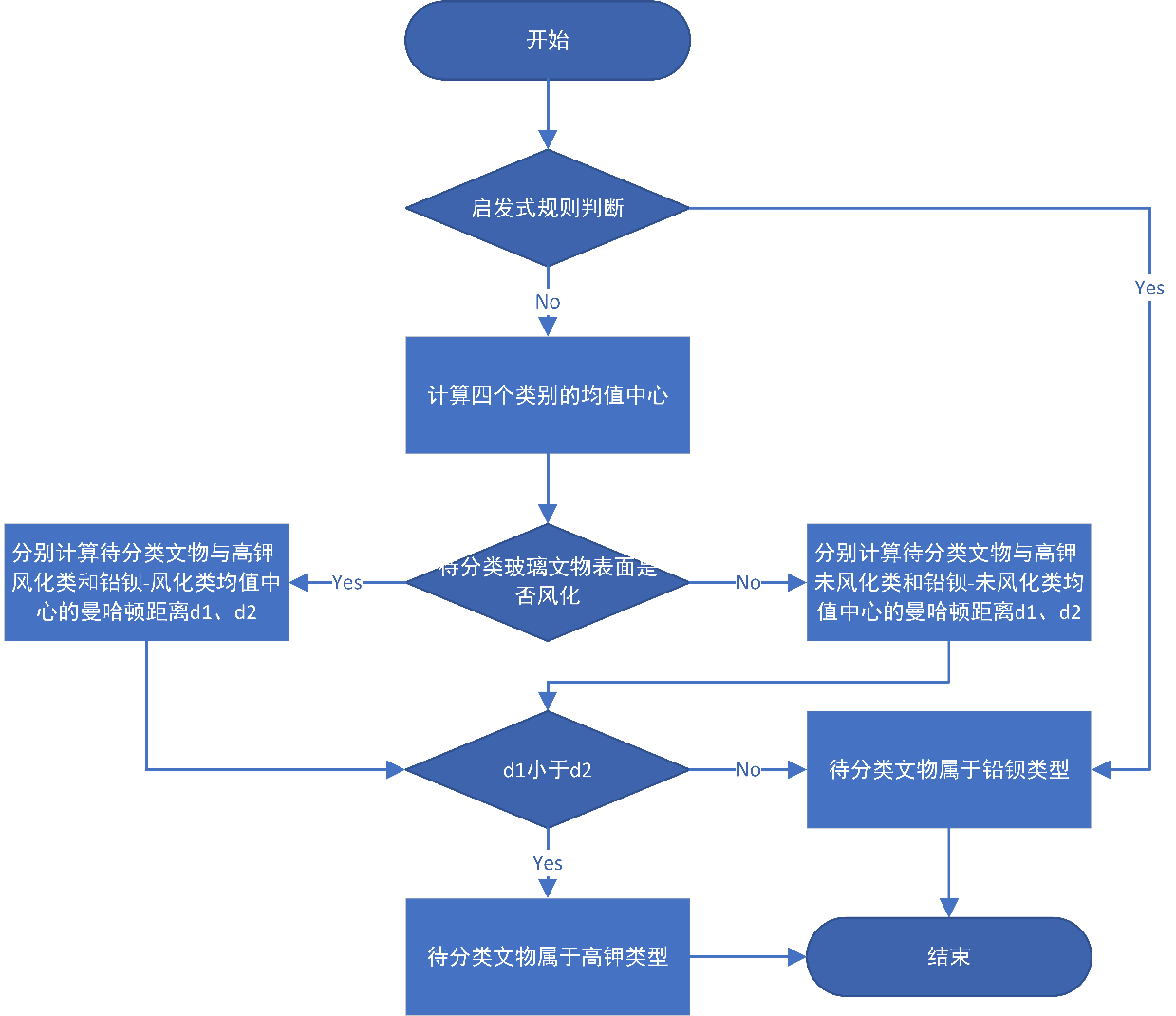


图 简单启发式均值分类模型流程图

使用简单启发式均值分类模型的详细步骤如下：

***Step*0**：启发式规则判断

由问题二的子问题1分析可得，铅钡玻璃文物的氧化铅含量较高，且高钾玻璃文物的氧化铅含量极少，故我们有一定理由相信具有较大氧化铅含量的文物属于铅钡文物。设题目所给的有效检测点数据中的铅钡文物检测点的最低氧化铅含量为，如果待分类玻璃文物的氧化铅含量大于等于，则我们直接认为其属于铅钡文物，程序结束；否则，进入Step1。

***Step*1**：求出每个类别的均值中心向量

设是第i个类别中的第j个数据（向量），则每个类别的均值中心向量的计算公式可以表达如下：

|  |  |
| --- | --- |
|  | () |

下图展示了二维平面上的均值中心，蓝色点为原始数据，红色点为蓝色点的均值中心：

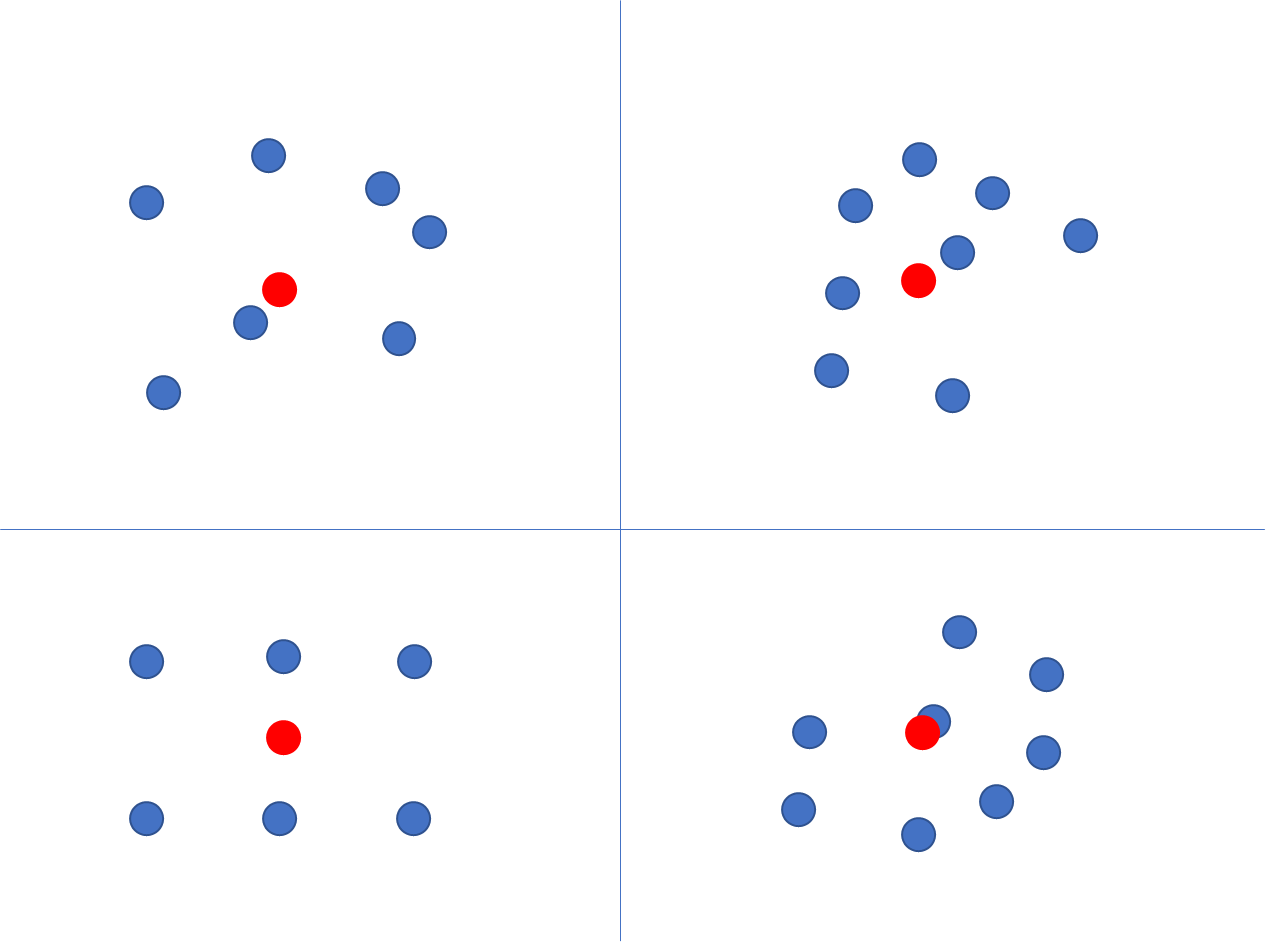


图 二维平面上的均值中心示意图

***Step*2**：计算待分类数据与其对应类型数据的均值中心的距离

设待分类数据的主成分特征数据向量为，为避免欧氏距离给小于1的正数据带来额外的缩小，我们采用曼哈顿距离对两个向量进行距离的评估，则待分类数据与对应类型数据的均值中心的距离计算公式如下：

|  |  |
| --- | --- |
|  | () |

***Step*3**：选出与待分类数据距离最小的均值中心所属的类别作为待分类数据的类别，程序结束

**例如**：假设高钾-风化数据的均值中心为（1，2），高钾-未风化数据的均值中心为（0，4），铅钡-风化的均值中心为（7，3），铅钡-未风化的均值中心为（2，1），待分类玻璃文物的数据向量为（2，3），已知其表面风化，假设其通过了启发式规则判断。

1. 由于已知待分类玻璃文物的表面风化，故其可能的分类属于高钾-风化和铅钡-风化类的其中一个
2. 分别计算待分类玻璃文物与高钾-风化类和铅钡-风化类均值中心的曼哈顿距离为：

|  |  |
| --- | --- |
|  | () |
|  | () |

1. 由于，故我们认为待分类文物属于高钾类型

## 分类结果

下表展示了所有待分类玻璃文物的文物编号、表面风化和使用简单启发式均值分类模型预测的类型信息。

表 待分类玻璃文物的分类结果表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **文物编号** | **表面风化** | **预测类型** | **氧化铅含量** | **D1** | **D2** |
| A1 | 无风化 | 高钾 | 0.00 | **14.45** | 49.91 |
| A2 | 风化 | 铅钡 | 34.3 | 38.36 | **21.80** |
| A3 | 无风化 | 铅钡 | 39.58 | 41.46 | **37.84** |
| A4 | 无风化 | 铅钡 | 24.28 | 43.35 | **16.13** |
| A5 | 风化 | 铅钡 | 12.23 | **24.62** | 48.57 |
| A6 | 风化 | 高钾 | 0 | **52.52** | 79.24 |
| A7 | 风化 | 高钾 | 0 | **48.00** | 76.84 |
| A8 | 无风化 | 铅钡 | 21.24 | 57.73 | **27.09** |

从上表可以看出，所有待分类玻璃文物均可通过启发式规则判断直接确定其所属类型。如果不考虑启发式规则判断，大部分分类结果不会改变，仅有文物编号为A5的待分类玻璃文物会因为D1<D2而被判定为高钾类型。

## 敏感性分析

为了验证简单启发式均值分类模型的敏感性，我们对题目所给的67个有效检测点数据进行了类型的预测，并绘制了相应的混淆矩阵（如下图所示），其中0代表高钾类型，1代表铅钡类型，所有有效检测点数据均预测正确，说明我们提出的简单启发式均值分类模型在小数据集上具有较好的分类效果。

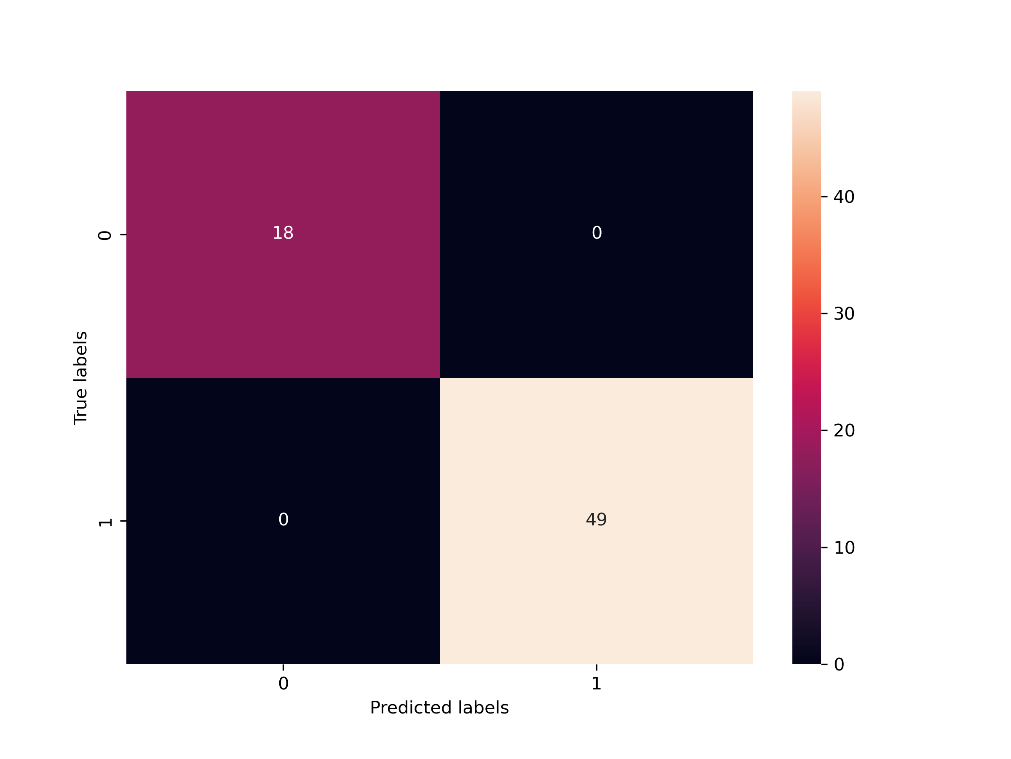


图 原始数据预测-混淆矩阵

# 模型的分析与检验

模型的分析与检验的内容也可以放到模型的建立与求解部分，这里我们单独抽出来进行讲解，因为这部分往往是论文的加分项，很多优秀论文也会单独抽出一节来对这个内容进行讨论。

模型的分析 ：在建模比赛中模型分析主要有两种，一个是灵敏度(性)分析，另一个是误差分析。灵敏度分析是研究与分析一个系统（或模型）的状态或输出变化对系统参数或周围条件变化的敏感程度的方法。其通用的步骤是：控制其他参数不变的情况下，改变模型中某个重要参数的值，然后观察模型的结果的变化情况。误差分析是指分析模型中的误差来源，或者估算模型中存在的误差，一般用于预测问题或者数值计算类问题。

模型的检验：模型检验可以分为两种，一种是使用模型之前应该进行的检验，例如层次分析法中一致性检验，灰色预测中的准指数规律的检验，这部分内容应该放在模型的建立部分；另一种是使用了模型后对模型的结果进行检验，数模中最常见的是稳定性检验，实际上这里的稳定性检验和前面的灵敏度分析非常类似，等会大家看到例子就明白了。

(大家尽量在论文中使用灵敏度分析，视频中有详细的讲解)

# 模型的评价、改进与推广

注：本部分的标题需要根据你的内容进行调整，例如：如果你没有写模型推广的话，就直接把标题写成模型的评价与改进。很多论文也把这部分的内容直接统称为“模型评价”部分，也是可以的。

## 模型的优点

优缺点是必须要写的内容，改进和推广是可选的，但还是建议大家写，实力比较强的建模者可以在这一块充分发挥，这部分对于整个论文的作用在于画龙点睛。

## 模型的缺点

缺点写的个数要比优点少

## 模型的改进

主要是针对模型中缺点有哪些可以改进的地方；

## 模型的推广

将原题的要求进行扩展，进一步讨论模型的实用性和可行性。

# 参考文献

所有引用他人或公开资料(包括网上资料)的成果必须按照科技论文的规范列出参考文献，并在正文引用处予以标注。

常见的三种参考文献的表达方式（标准不唯一）：

书籍的表述方式为： [编号] 作者，书名，出版地：出版社，出版年月。

期刊杂志论文的表述方式为： [编号] 作者，论文名，杂志名，卷期号：起止页码，出版年。

网上资源(例如数据库、政府报告)的表述方式为： [编号] 作者，资源标题，网址，访问时间。

附录

|  |
| --- |
| 附录1 |
| 介绍：支撑材料的文件列表 |
| 这是最近国赛要求加入的一个部分，大家可以看我讲的论文写作视频。  <https://www.bilibili.com/video/BV1Na411w7c2> |

|  |
| --- |
| 附录2 |
| 介绍：该代码是某某语言编写的，作用是什么 |
|  |

|  |
| --- |
| 附录3 |
| 介绍：该代码是某某语言编写的，作用是什么 |
|  |

除了支撑材料的文件列表和源程序代码外，附录中还可以包括下面内容：

* 某一问题的详细证明或求解过程；
* 自己在网上找到的数据；
* 比较大的流程图；
* 较繁杂的图表或计算结果

1. 王金桃, 周利锋, 高尔生. 第六讲 卡方检验[J]. 实验动物与比较医学, 2000(4):251-254. [↑](#footnote-ref-1)
2. 耿丽娟,李星毅.用于大数据分类的KNN算法研究[J].计算机应用研究,2014,31(05):1342-1344+1373. [↑](#footnote-ref-2)