

# 武汉大学计算机学院

## 本科生课程设计报告

### 分类分析实验

专 业 名 称: 弘毅学堂计算机科学与技术

课 程 名 称: 数据科学导论

指 导 教 师 一: 吴小莹 教授

学 生 学 号: 2018300003088

学 生 姓 名: 刘阳飞晓

二〇二一年四月

# 郑 重 声 明

本人呈交的设计报告，是在指导老师的指导下，独立进行实验工作所取得的成果，所有数据、图片资料真实可靠。尽我所知，除文中已经注明引用的内容外，本设计报告不包含他人享有著作权的内容。对本设计报告做出贡献的其他个人和集体，均已在文中以明确的方式标明。本设计报告的知识产权归属于培养单位。

本人签名: \_\_\_\_\_

日期: 2021.4.16

## 摘 要

本文主要介绍和讨论了 ID3 算法和朴素贝叶斯算法之间对于 iris 数据集的分类结果以及两种算法之间的差距

**关键词:** Weka; 分类分析; ID3; 朴素贝叶斯算法

# 目 录

<b>1 先说重要的</b>	<b>1</b>
1.1 具体使用步骤 . . . . .	1
1.2 编译的方法 . . . . .	1
1.3 文档类型选择 . . . . .	1
1.4 打印的问题 . . . . .	1
<b>2 杂七杂八的话</b>	<b>3</b>
2.1 Readme . . . . .	3
2.2 字体调节 . . . . .	3
2.3 字号调节 . . . . .	3
2.4 已加入的常用宏包 . . . . .	4
2.5 标点符号的问题 . . . . .	5
2.6 引用的问题 . . . . .	5
2.6.1 参考文献的引用 . . . . .	5
2.6.2 定理和公式的引用 . . . . .	5
2.7 图形与表格 . . . . .	6
<b>3 其他事项</b>	<b>8</b>
<b>结论</b>	<b>9</b>
<b>参考文献</b>	<b>10</b>
<b>附录 A 测试</b>	<b>11</b>
A.1 第一个测试 . . . . .	11



# 1 先说重要的

## 1.1 具体使用步骤

**Step 1** 进入 includefile 文件夹, 打开 frontmatter.tex, backmatter.tex 这两个文档, 分别填写 (1) 中文摘要, (2) 实验结论.

**Step 2** 打开主文档 Experiment-template.tex, 填写题目、学生姓名等信息, 书写正文.

**Step 3** 使用 XeLaTeX 编译. 具体见 1.2 节.

## 1.2 编译的方法

默认使用 XeLaTeX 编译, 直接生成 pdf 文件.

若另存为新文档, 请确保文档保存类型为 :UTF-8. 当然目前很多编辑器默认文字编码为 UTF-8. WinEdt 9.0 之后的版本都是默认保存为 UTF-8 的.

## 1.3 文档类型选择

**本小节是毕业论文打印介绍, 实验报告可以略过**

文档类型有 2 种情形:

---

<code>\documentclass{WHUBachelor}</code>	毕业论文
<code>\documentclass[forprint]{WHUBachelor}</code>	毕业论文打印版

---

相关解释见下节.

## 1.4 打印的问题

**本小节是毕业论文打印介绍, 实验报告可以略过**

- i) 关于文档选项 forprint: 交付打印时, 建议加上选项 forprint, 以消除链接文字之彩色, 避免打印字迹偏淡.
- ii) 打印时留意不要缩小页面或居中. 即页面放缩方式应该是“无”(Adobe Reader XI 是选择“实际大小”). 有可能页面放缩方式默认为“适合可打印区域”, 会导致打印为原页面大小的 97%. 文字不要居中打印, 是因为考虑到装订, 左侧的空白留得稍多一点 (模板已作预留).
- iii) 遗留问题: 封面需要打印部重新制作. 校内打印部通常有现成的模板. 我们自己做封面, 打印部不一定好用.

问: 生成 PDF 文件时, 不能去掉目录和文章的引用彩色方框, 请问怎么解决?

答: 方框表示超级链接, 只在电脑上看得见. 实际打印时, 是没有的. 另外, 文档类型加选项 forprint 之后, 这些框框会隐掉的.

本文档下载更新地址: <https://github.com/xiaoxinganling/WHUExperiment>.  
使用之前, 请移步查看是否有更新.

问题反馈及建议, 请联系: mxzhou1998@gmail.com.

## 2 杂七杂八的话

### 2.1 Readme

模板文件的结构, 如下表所示:

Experiment-template.tex		主文档. 在其中填写正文.
includefile 文件夹	frontmatter.tex	郑重声明、摘要.
	backmatter.tex	实验结论.
figures 文件夹		存放图片文件.
WHUBachelor.cls		定义文档格式的 class file. 不可删除.

无需也不要改变、移动上述文档的位置.

如果不习惯用 `\include{ }` 的方式加入“子文档”, 当然可以把它们合并在主文档, 成为一个文档. (但是这样并不会给我们带来方便.)

### 2.2 字体调节

`\songti`    宋体  
`\heiti`     黑体  
`\fangsong` 仿宋  
`\kaishu`    楷书

### 2.3 字号调节

字号命令: `\zihao`



<code>\zihao{0}</code>	初号字 English
<code>\zihao{-0}</code>	小初号 English
<code>\zihao{1}</code>	一号字 English
<code>\zihao{-1}</code>	小一号 English
<code>\zihao{2}</code>	二号字 English
<code>\zihao{-2}</code>	小二号 English
<code>\zihao{3}</code>	三号字 English
<code>\zihao{-3}</code>	小三号 English
<code>\zihao{4}</code>	四号字 English
<code>\zihao{-4}</code>	小四号 English
<code>\zihao{5}</code>	五号字 English
<code>\zihao{-5}</code>	小五号 English
<code>\zihao{6}</code>	六号字 English
<code>\zihao{-6}</code>	小六号 English
<code>\zihao{7}</code>	七号字 English
<code>\zihao{8}</code>	八号字 English

## 2.4 已加入的常用宏包

**cite** 参考文献引用, 得到形如 [3-7] 的样式.

**color,xcolor** 支持彩色.

**enumerate** 方便自由选择 enumerate 环境的编号方式. 比如

`\begin{enumerate}[(a)]` 得到形如 (a), (b), (c) 的编号.

`\begin{enumerate}[i]` 得到形如 i), ii), iii) 的编号.

`\begin{enumerate}[\hspace{1cm}(1)]` `\hspace`命令用于调整距离

另外要说明的是, itemize, enumerate, description 这三种 list 环境, 已经调节了其间距和缩进, 以符合中文书写的习惯.

## 2.5 标点符号的问题

建议使用半角的标点符号, 后边再键入一个空格. 特别是在英文书写中要注意此问题!

双引号是由两个左单引号、两个右单引号构成的: `` `'. 左单引号在键盘上数字 1 的左边.

但是, 无论您偏向于全角或半角, 强烈建议您使用实心的句号, 只要您书写的是自然科学的文章. 原因可能是因为, 比如使用全角句号的句子结尾处的“ $x$ 。”容易误为数学式  $x_0$ (\$x\_0\$) 吧.

## 2.6 引用的问题

### 2.6.1 参考文献的引用

参考文献的引用, 用命令 `\cite{ }`. 大括号内要填入的字串, 是自命名的文献条目名.

比如, 通常我们会说:

关于此问题, 请参见文献 [3]. 作者某某还提到了某某概念<sup>[2]</sup>.

上文使用的源文件为:

关于此问题, 请参见文献 `\cite{r2}`. 作者某某还提到了某某概念 `\upcite{r1}`.

其中 `\upcite` 是自定义命令, 使文献引用呈现为上标形式.

(注意: 这里文献的引用, 有时需要以上标形式出现, 有时需要作为正文文字出现, 为什么?)

另外, 要得到形如 [2, 4, 5, 6] 的参考文献连续引用, 需要用到 `cite` 宏包 (模板已经加入), 在正文中使用 `\cite{r1,r3,r4,r5}` 的引用形式即可. 或者, 连续引用的上标形式: 使用 `\upcite{r1,r2,r3}`, 得到<sup>[2, 3, 4]</sup>.

### 2.6.2 定理和公式的引用

**定理 2.6.1** (谁发现的) 最大的正整数是 1.

**证明** 要找到这个最大的正整数, 我们设最大的正整数为  $x$ , 则  $x \geq 1$ , 两边

同时乘以  $x$ , 得到

$$x^2 \geq x. \quad (2.1)$$

而  $x$  是最大的正整数, 由 (2.1) 式得到

$$x^2 = x.$$

所以

$$x = 1. \quad \square$$

定理 2.6.1 是一个重大的发现.

**定义 2.6.1 (整数)** 正整数 (例如 1, 2, 3)、负整数 (例如  $-1$ ,  $-2$ ,  $-3$ ) 与零 (0) 合起来统称为**整数**.

**注 2.6.1** 整数集合在数学上通常表示为  $\mathbf{Z}$  或  $\mathbb{Z}$ , 该记号源于德语单词 Zahlen(意为“数”)的首字母.

**性质 2.6.1** 任意两个整数相加、相减、相乘的结果, 仍然是整数.

**例 2.6.1**  $1 + 2 = 3$ .

**推论 2.6.1** 在整数集合内, 相加、相减、相乘运算是封闭的.

## 2.7 图形与表格

支持对 eps, pdf, jpg 等等常见图形格式.

再次 **澄清一个误会**: L<sup>A</sup>T<sub>E</sub>X 支持的图形格式绝非 eps 这一种. 无需特意把图片转化为 eps.

用形如 `\includegraphics[width=12cm]{Daisy.jpg}` 的命令可以纳入图片.

如图 2.1 是一个纳入 jpg 图片的例子.

表格问题, 建议使用“三线表”, 如表 2.1.



图 2.1 一个彩色 jpg 图片的例子

表 2.1 一般三线表

123	4	5	123	4	5123	4	5	123	4	5
67	890	13	123	4	5123	4	5	123	4	5
67	890	13	123	4	5123	4	5	123	4	5
67	890	13	123	4	5123	4	5	123	4	5

### 3 其他事项

以下是广告时间, 插播一段广告:

- 插图的制作, 建议用 pgf, 也叫 tikz. pgf 的长处是源文件直接植入 T<sub>E</sub>X 文档, 管理起来非常方便. 这里有我写的一个关于初次使用 pgf 的帖子:

<http://bbs.ctex.org/forum.php?mod=viewthread&tid=30480>.

- 生成参考文献, 建议使用 BibTeX.:

<http://bbs.ctex.org/forum.php?mod=viewthread&tid=26056>.

使用 BibTeX 做参考文献时, 借助 EndNote 或者 NoteExpress, 可以非常漂亮简单地解决 bib 文件的录入问题. NoteExpress 在校图书馆网站有正版软件提供下载. 当然 EndNote 本身就是 Thomson Corporation 推出的 (和 SCI 搜索引擎是同一家公司), 和多个重要文献搜索引擎有良好的功能配合.

Google 学术搜索也提供了文献的 bib 格式. 录入参考文献时, 偶尔用一用 Google 学术搜索, 还可以核查或减少录入的错误, 并减少录入的工作量.

- 幻灯片的制作, 建议使用 Beamer. 这里有我写的一个模板, 谨供参考:

<http://bbs.ctex.org/forum.php?mod=viewthread&tid=27695>.

## 结    论

这里写本次实验的结论。

## 参考文献

- [1] Dean J, Ghemawat S. MapReduce: Simplified Data Processing on Large Clusters[A].Eric A. Brewer, Peter Chen.6th Symposium on Operating Systems Design and Implementation(OSDI 2004)[C], San Francisco, California, USA: USENIX Association, 2004:137–150.
- [2] 作者. 文章题目 [J]. 期刊名, 出版年份, 卷号 (期数): 起止页码.
- [3] 作者. 书名 [M]. 版次. 出版地: 出版单位, 出版年份: 起止页码.
- [4] 邓建松等, 《L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> 科技排版指南》, 科学出版社.
- [5] 吴凌云, 《C<sub>T</sub>E<sub>X</sub> FAQ (常见问题集)》, *Version 0.4*, June 21, 2004.
- [6] Herbert Voß, Mathmode, <http://www.tex.ac.uk/ctan/info/math/voss/mathmode/Mathmode.pdf>.

## 附录 A 测试

### A.1 第一个测试

测试公式编号

$$1 + 1 = 2. \tag{A.1}$$

表格编号测试

表 A.1 测试表格

11	13	13	13	13
12	14	13	13	13



## 附录 B 附录测试

## 教师评语评分

评语:

评分:

评阅人:

年 月 日

(备注:对该实验报告给予优点和不足的评价,并给出百分制评分。)