

Preface

前言

感谢

首先感谢大家的信任。

作者仅仅是在学习应用数学科学和机器学习算法时，多读了几本数学书，多做了些思考和知识整理而已。知者不言，言者不知。知者不博，博者不知。水平有限，把自己有限所学所思斗胆和大家分享，作者权当无知者无畏。希望大家在 B 站视频下方和 Github 多提意见，让这套书成为作者和读者共同参与创作的优质作品。

特别感谢清华大学出版社的栾大成老师。从选题策划、内容创作、装帧设计，栾老师事无巨细、一路陪伴。每次和栾老师交流，我都能感受到他对优质作品的追求、对知识分享的热情。

出来混总是要还的

曾几何时，考试是我们学习数学的唯一动力。考试是头悬梁的绳，是锥刺股的锥。我们中的绝大多数人从小到大为各种考试埋头题海，数学味同嚼蜡，甚至让人恨之入骨。

数学给我们带来了无尽的折磨。我们憎恨数学，恐惧数学，恨不得一走出校门就把数学抛之脑后、老死不相往来。

可悲可笑的是，我们其中很多人可能会在毕业的五年或十年以后，因为工作需要，不得不重新学习微积分、线性代数、概率统计，悔恨当初没有学好数学、走了很多弯路、没能学以致用，从而迁怒于教材和老师。

这一切不能都怪数学，值得反思的是我们学习数学的方法、目的。

再给自己一个学数学的理由

为考试而学数学，是被逼无奈的举动。而为数学而数学，则又太过高尚而遥不可及。

相信对于绝大部分的我们来说，数学是工具、是谋生手段，而不是目的。我们主动学数学，是想用数学工具解决具体问题。

现在，这套书给大家一个“学数学、用数学”的全新动力——数据科学、机器学习。

数据科学和机器学习已经深度融合到我们生活的方方面面，而数学正是开启未来大门的钥匙。不是所有人生来都握有一副好牌，但是掌握“数学 + 编程 + 机器学习”绝对是王牌。这次，学习数学不再是为了考试、分数、升学，而是投资时间、自我实现、面向未来。

未来已来，你来不来？

本套丛书如何帮到你

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

为了让大家学数学、用数学，甚至爱上数学，作者可谓颇费心机。在创作这套书时，作者尽量克服传统数学教材的各种弊端，让大家学习时有兴趣、看得懂、有思考、更自信、用得着。

为此，丛书在内容创作上突出以下几个特点：

- ◀ **数学 + 艺术**——全彩图解，极致可视化，让数学思想跃然纸上、生动有趣、一看就懂，同时提高大家的数据思维、几何想象力、艺术感；
- ◀ **零基础**——从零开始学习 Python 编程，从写第一行代码到搭建数据科学和机器学习应用；
- ◀ **知识网络**——打破数学板块之间的壁垒，让大家看到数学代数、几何、线性代数、微积分、概率统计等板块之间的联系，编织一张绵密的数学知识网络；
- ◀ **动手**——授人以鱼不如授人以渔，和大家一起写代码、用 Streamlit 创作数学动画、交互 App；
- ◀ **学习生态**——构造自主探究式学习生态环境“微课视频 + 纸质图书 + 电子图书 + 代码文件 + 可视化工具 + 思维导图”，提供各种优质学习资源；
- ◀ **理论 + 实践**——从加减乘除到机器学习，丛书内容安排由浅入深、螺旋上升，兼顾理论和实践；在编程中学习数学，学习数学时解决实际问题。

虽然本书标榜“从加减乘除到机器学习”，但是建议读者朋友们至少具备高中数学知识。如果读者正在学习或曾经学过大学数学（微积分、线性代数、概率统计），这套书就更容易读了。

聊聊数学

数学是工具。锤子是工具，剪刀是工具，数学也是工具。

数学是思想。数学是人类思想的高度抽象的结晶体。在其冷酷的外表之下，数学的内核实际上就是人类朴素的思想。学习数学时，知其然，更要知其所以然。不要死记硬背公式定理，理解背后的数学思想才是关键。如果你能画一幅图、用大白话描述清楚一个公式、一则定理，这就说明你真正理解了它。

数学是语言。就好比世界各地不同种族有自己的语言，数学则是人类共同的语言和逻辑。数学这门语言极其精准、高度抽象，放之四海而皆准。虽然我们中绝大多数人没有被数学女神选中，不能为人类的对数学认知开疆扩土；但是，这丝毫不妨碍我们使用数学这门语言。就好比，我们不会成为语言学家，我们完全可以使用母语和外语交流。

数学是体系。代数、几何、线性代数、微积分、概率统计、优化方法等等，看似一个个孤岛，实际上都是数学网络的一条条织线。建议大家学习时，特别关注不同数学板块之间的联系，见树，更要见林。

数学是基石。拿破仑曾说“数学的日臻完善和这个国强民富息息相关。”数学是科学进步的根基，是经济繁荣的支柱，是保家卫国的武器，是探索星辰大海的航船。

数学是艺术。数学和音乐、绘画、建筑一样，都是人类艺术体验。通过可视化工具，我们会在看似枯燥的公式、定理、数据背后，发现数学之美。

数学是历史，是人类共同记忆体。“历史是过去，又属于现在，同时在指引未来。”数学是人类的集体学习思考，她把人的思维符号化、形式化，进而记录、积累、传播、创新、发展。从甲

骨、泥板、石板、竹简、木牍、纸草、羊皮卷、活字印刷、纸质书，到数字媒介，这一过程持续了数千年，至今绵延不息。

数学是无穷无尽的**想象力**，是人类的**好奇心**，是自我挑战的**毅力**，是一个接着一个的**问题**，是看似荒诞不经的**猜想**，是一次次胆大包天的**批判性思考**，是敢于站在前人的臂膀之上的**勇气**，是孜孜不倦地延展人类认知边界的**不懈努力**。

家园、诗、远方

诺瓦利斯曾说：“哲学就是怀着一种乡愁的冲动到处去寻找家园。”

在纷繁复杂的尘世，数学纯粹的就像精神的世外桃源。数学是，一束光，一条巷，一团不灭的希望，一股磅礴的力量，一个值得寄托的避风港。

打破陈腐的锁链，把功利心暂放一边，我们一道怀揣一分乡愁、心存些许诗意、踩着艺术维度，投入数学张开的臂膀，驶入她色彩斑斓、变幻无穷的深港，感受久违的归属，一睹更美、更好的远方。

Acknowledgement

致谢

To my parents.

谨以此书献给我的母亲父亲

How to Use the Book

使用本书

丛书资源

本系列丛书提供的配套资源有以下几个：

- 纸质图书；
- PDF 文件，方便移动终端学习；请大家注意，纸质图书经过出版社五审五校修改，内容细节上会和 PDF 文件有出入。
- 每章提供思维导图，纸质书提供全书思维导图海报；
- Python 代码文件，直接下载运行，或者复制、粘贴到 Jupyter 运行；
- Python 代码中有专门用 Streamlit 开发数学动画和交互 App 的文件；
- 微课视频，强调重点、讲解难点、聊聊天。

在纸质书中为了方便大家查找不同配套资源，作者特别设计了如下几个标识。



数学家、科学家、
艺术家等语录



代码中核心Python
库函数和讲解



思维导图总结本章
脉络和核心内容



配套Python代码完
成核心计算和制图



用Streamlit开发制
作App应用



介绍数学工具、机
器学习之间联系



引出本书或本系列
其他图书相关内容



提醒读者格外注意
的知识点



每章配套微课视频
二维码



相关数学家生平贡
献介绍



每章结束总结或升
华本章内容



本书核心参考和推
荐阅读文献

微课视频

本书配套微课视频均发布在 B 站——生姜 DrGinger：

◀ <https://space.bilibili.com/513194466>

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

微课视频是以“聊天”的方式，和大家探讨某个数学话题的重点内容，讲讲代码中可能遇到的难点，甚至侃侃历史、说说时事、聊聊生活。

本书配套的微课视频目的是引导大家自主编程实践、探究式学习，并不是“照本宣科”。

纸质图书上已经写得很清楚的内容，视频课程只会强调重点。需要说明的是，图书内容不是视频的“逐字稿”。

代码文件

本系列丛书的 Python 代码文件下载地址为：

◀ <https://github.com/Visualize-ML>

Python 代码文件会不定期修改，请大家注意更新。图书配套的 PDF 文件和勘误也会上传到这个 GitHub 账户。因此，建议大家注册 GitHub 账户，给书稿文件夹标星 (star) 或分支克隆 (fork)。

考虑再三，作者还是决定不把代码全文印在纸质书中，以便减少篇幅，节约用纸。

本书编程实践例子中主要使用“鸢尾花数据集”，数据来源是 Scikit-learn 库、Seaborn 库。此外，系列丛书封面设计致敬梵高《鸢尾花》，要是给本系列丛书起个昵称的话，作者乐见“鸢尾花书”。

App 开发

本书几乎每一章都至少有一个用 Streamlit 开发的 App，用来展示数学动画、数据分析、机器学习算法。

Streamlit 是个开源的 Python 库，能够方便快捷搭建、部署交互型网页 App。Streamlit 非常简单易用、很受欢迎。Streamlit 兼容目前主流的 Python 数据分析库，比如 NumPy、Pandas、Scikit-learn、PyTorch、TensorFlow 等等。Streamlit 还支持 Plotly、Bokeh、Altair 等交互可视化库。

本书中很多 App 设计都采用 Streamlit + Plotly 方案。此外，本书专门配套教学视频手把手和大家一起做 App。

大家可以参考如下页面，更多了解 Streamlit：

◀ <https://streamlit.io/gallery>

◀ <https://docs.streamlit.io/library/api-reference>

实践平台

本书作者编写代码时采用的 IDE (integrated development environment) 是 Spyder，目的是给大家提供简洁的 Python 代码文件。

但是，建议大家采用 JupyterLab 或 Jupyter notebook 作为本系列丛书配套学习工具。

简单来说，Jupyter 集合“浏览器 + 编程 + 文档 + 绘图 + 多媒体 + 发布”众多功能与一身，非常适合探究式学习。

运行 Jupyter 无需 IDE，只需要浏览器。Jupyter 容易分块执行代码。Jupyter 支持 inline 打印结果，直接将结果图片打印在分块代码下方。Jupyter 还支持很多其他语言，比如 R 和 Julia。

使用 markdown 文档编辑功能，可以编程同时写笔记，不需要额外创建文档。Jupyter 中插入图片和视频链接都很方便。此外，还可以插入 Latex 公式。对于长文档，可以用边栏目录查找特定内容。

Jupyter 发布功能很友好，方便打印成 HTML、PDF 等格式文件。

Jupyter 也并不完美，目前尚待解决的问题有几个。Jupyter 中代码调试不方便，需要安装专门插件 (比如 debugger)。Jupyter 没有 variable explorer，要么 inline 打印数据，要么将数据写到 csv 或 Excel 文件中再打开。图像结果不具有交互性，比如不能查看某个点的值，或者旋转 3D 图形，可以考虑安装 (jupyter-matplotlib)。注意，利用 Altair 或 Plotly 绘制的图像支持交互功能。对于自定义函数，目前没有快捷键直接跳转到其定义。但是，很多开发者针对这些问题都开发了插件，请大家留意。

大家可以下载安装 Anaconda，JupyterLab、Spyder、PyCharm 等常用工具都集成在 Anaconda 中。下载 Anaconda 的地址为：

◀ <https://www.anaconda.com/>

学习步骤

大家可以根据自己的偏好制定学习步骤，本书推荐如下步骤。



学完每章后，大家可以在平台上发布自己的 Jupyter 笔记，进一步听取朋友们的意见，共同进步。这样做还可以提高自己学习的动力。

意见建议

欢迎大家对本系列丛书提意见和建议，丛书专属邮箱地址为：

◀ jiang.visualize.ml@gmail.com

也欢迎大家在 B 站视频下方留言互动。

Contents

目录



Introduction

绪论

图解 + 编程 + 实践 + 数学板块融合 + 历史 + 英文术语

0.1 本册在全套丛书的定位

本系列丛书有三大板块——编程、数学、实践。机器学习各种算法离不开数学，而《数学要素》一册是“数学”板块的第一册。本书介绍的数学工具是整个“数学”板块的基础，当然也是数据科学和机器学习实践的基础。

《数学要素》一册中编程和可视化无处不在，限于篇幅本书不会专门讲解编程基础内容。因此，建议编程零基础读者先学习《编程不难》和《可视之美》两册内容。当然，根据个人情况，平行学习《数学要素》、《编程不难》和《可视之美》，也是可行的。

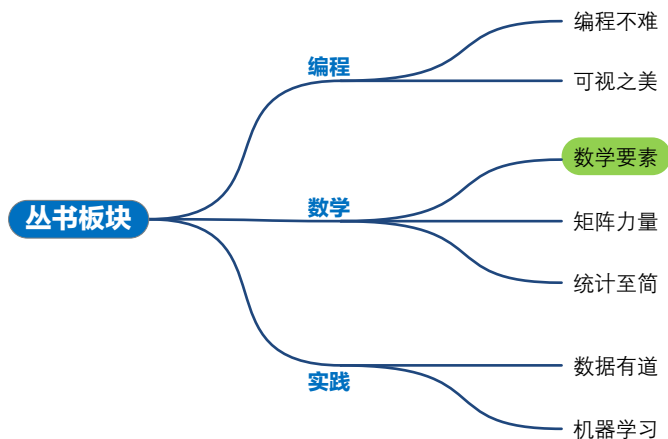


图 1. 本系列丛书板块布局

0.2 结构：7 大板块

本书可以归纳为 7 大板块——基础、坐标系、函数、解析几何、微积分、概率统计、线性代数。

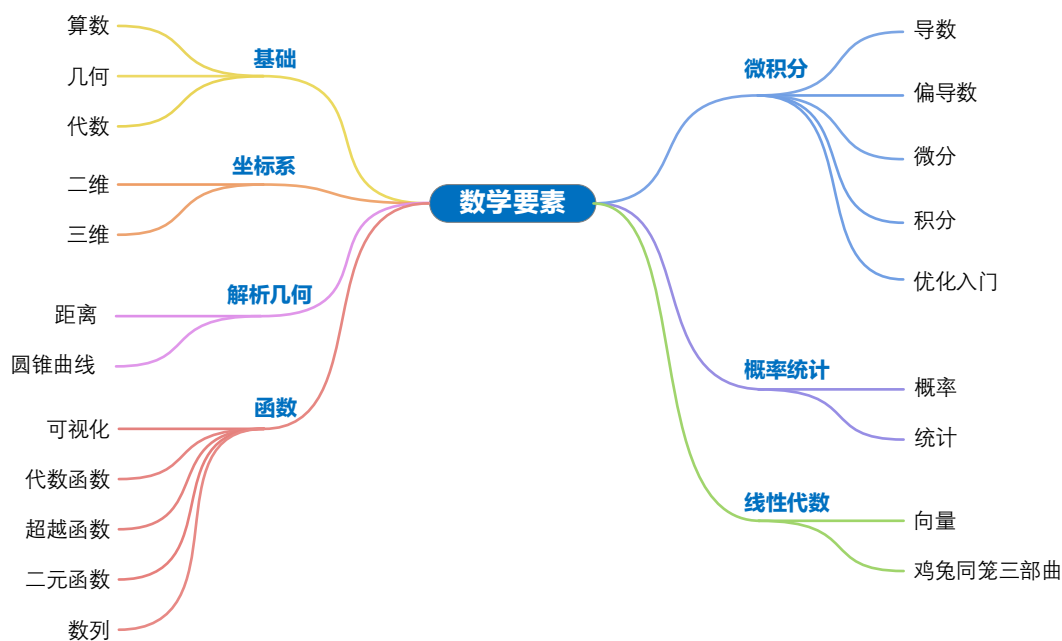


图 2. 《数学要素》板块布局

基础部分

基础部分从加、减、乘、除四则运算讲起。线性代数在机器学习中应用广泛，本书第 1、2 章开门见山讲向量和矩阵的基本运算，也会在本册各个板块见缝插针介绍线性代数基础知识。

本书第 3 章回顾常用几何知识，几何视角是本系列丛书的一大特色。这一章有一大亮点——圆周率估算。圆周率估算是本书的一条重要线索，本书会按时间先后顺序介绍如何用不同数学工具估算圆周率。

第 4 章回顾代数知识，里面有两个亮点值得大家特别注意：一个是杨辉三角，本书后面会将杨辉三角和概率统计、随机过程联系起来；另一个是鸡兔同笼问题，本书最后三章都围绕鸡兔同笼这个话题展开。

坐标系

笛卡尔坐标系让几何和代数走到一起，本书第 5、6 两章介绍坐标系有关内容。这两章的一大特色是——代数式可视化，几何体参数化。没有坐标系，就没有函数，也不会有微积分；因此，坐标系的地位毋庸置疑。

解析几何

第 7、8、9 三章介绍解析几何内容，其中有两大亮点——距离度量、椭圆。距离度量中，大家要善于用等距线这个可视化工具。此外，大家需要注意欧氏距离并不是唯一的距离度量。第二个亮点是椭圆，椭圆可谓“多面手”，大家很快会看到椭圆在概率统计、线性代数、数学科学、机器学习中大放异彩。

函数

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

第 10 章到第 14 章都是围绕函数展开。有几点值得强调，学习任何函数时，建议大家编程绘制函数线图，以便观察函数形状、变化趋势。此外，学会利用曲面、剖面线、等高线等可视化工具观察分析二元函数。再者，不同函数都有自身特定性质，对应独特应用场景。第 14 章讲解数列，数列可以看成是特殊的函数。本章中，累加、极限这两个知识点特别值得关注，它们都是微积分基础。

微积分

第 15 章到第 19 章讲解微积分以及优化问题内容。牛顿和莱布尼兹分别发明微积分之后，整个数学王国的版图天翻地覆。导数、偏导数、微分、积分给我们提供研究函数性质的量化工具。学好这四章的秘诀就是——几何图解。导数是切线斜率，偏导数是某个变量方向上切线斜率，微分是线性近似，泰勒展开是多项式函数叠加，积分是求面积，二重积分是求体积。数据科学、机器学习中所有算法都可以写成优化问题，而构造、求解优化问题离不开微积分。因此，本书在讲完微积分之后立刻安排了第 19 章，介绍优化问题入门知识。本系列丛书后续还会在各册不断介绍优化方法。

概率统计

第 20、21 两章是概率统计入门。本系列丛书专门有《概率统计》一册系统讲解这个版块，但是这并不意味着本书第 20、21 两章内容毫无出彩之处；相反，这两章亮点颇多。第 20 章概率内容实际上是代数部分杨辉三角的延伸，本章用二叉树这个知识点，将代数和概率统计串联在一起。第 20 章最后还介绍了随机过程。第 21 章的关键词就是“图解”，用图像可视化数据，用图像展示概率统计定义。

线性代数

本书最后四章以线性代数收尾。第 22 章可视化向量和向量运算。第 23、24、25 三章是“鸡兔同笼三部曲”，这三章虚构了一个世外桃源，讲述与世隔绝的村民如何利用舶来的线性代数知识，解决村民养鸡养兔时遇到的数学疑难杂症。这三章涉及线性方程组、向量空间、投影、最小二乘法线性回归、马尔科夫过程、特征值分解等内容。这三章一方面给大家展示本书重要数学工具的应用，另外这三章也为本系列丛书《矩阵力量》一册做了内容预告和铺垫。

0.3 特点：知识融合

《数学要素》打破数学板块的藩篱，将算数、代数、线性代数、几何、解析几何、概率统计、微积分、优化方法等板块有机结合在一起。

作为丛书的核心特点，《数学要素》一册内容编排上突出“图解 + 编程 + 机器学习应用”。讲解一些特定数学工具时，本书会穿插介绍其在数据科学和机器学习领域应用场景，让大家学以致用。本书主要使用 Python 的这几个库——Scipy、Numpy、SymPy、Matplotlib、Seaborn 等。

《数学要素》一册还强调数学文化，内容安排上尽可能沿着数学发展先后脉络，为大家展现整副历史图景。本书还介绍数学史上关键人物，让大家看到数学如何薪火相传、接续发展。

为了帮助大家阅读英文文献以及学术交流，本书还特别总结常用数学知识的英文表述。

下面让我们一起开始《数学要素》一册的学习之旅。