

- [关于本书](#)
- [前置基础](#)
- [AI、ML、DL都是啥](#)
- [参考资料](#)
- [学习建议](#)
- [优雅快捷的写公式-latex](#)
- [环境搭建](#)
- [第一章补充说明](#)

## 1 关于本书

### About this book

首先感谢亲爱的读者朋友们抽出宝贵的时间阅读本书。本书是【AI精研社】零基础入门AI解决方案(以下简称本方案)的一部分。

本书是【AI精研社】以特有的方式(极简体验)对李航老师所著《统计学习方法(第2版)》的解读。本书的章节并非直接对应原书，是专为初学者重新设计的学习路径，但原书的重要基础部分，本书及后续系列都会以生动、易懂、详尽的方式解读给同学们。

本方案旨在宣传 learning by doing, understanding by creating, evolving from the simplest 的极简体验学习倡议。

本方案尝试帮助同学们在入门AI的过程中，平滑顺畅的从理论联系到实战，将数学符号转换成Python代码，将数学符号、Python代码、理论概念一一对应。

本方案以 Python 编程语言为基础，通过极少量(如3行代码)极易懂的代码片段、运行过程、运行结果，对概念、算法进行解释说明。

## 2 前置基础

### Prerequisites

因此，本书假定读者朋友们具备少量的Python编程经验(20小时、200行以上的习题、项目经验)及相关常识。具体包含: Python基础语法、Python包管理、NumPy、Matplotlib、Jupyter Notebook、Google Colab、Binder。

如果需要了解上述内容，可以在手机应用"微信读书"上搜索《人工智能极简编程入门》，免费获取【AI精研社】专为入门ML而优化的Python基础教程。

本书假定同学们对人工智能已有简单了解，知道人工智能可以应用在哪些场景(部分)及其效果，从而建立起对人工智能的直观感受。

尚不了解人工智能的应用情况，对人工智能没有常识性了解的同学建议先从央视《机智过人》开始，获取初步体验。

[机智过人 央视官网 \(http://tv.cctv.com/lm/jzgr/\)](http://tv.cctv.com/lm/jzgr/)

[机智过人 youtube频道 \(https://www.youtube.com/playlist?list=PL0eGJygpmOH6ywJc2QueT6mIBs9yrmnZ-\)](https://www.youtube.com/playlist?list=PL0eGJygpmOH6ywJc2QueT6mIBs9yrmnZ-)

除此之外，还给童鞋们推荐一部美剧《数字追凶》(NUMB3RS)。在培养兴趣、了解AI算法与数学基础的同时还能练习英语。而英语、数学是入门AI算法无法绕过的必备基础。

在观赏电视节目的同时，同学们请思考并尝试回答以下问题

1. 人工智能长什么样，有什么用？
2. 人工智能能做什么、不能做什么？

### 3 AI、ML、DL都是啥

AI ML DL

如果被突如其来的英语吓到了，可以先看本小节末尾的总结-:)

在对人工智能建立起直观感受之后，同学们请深吸一口气，准备开始专业之旅。

即使是入门，也是在尝试进入一个领域，一个行当。而进入一个新领域时，往往要学习一系列的行话、术语。

入门机器学习也同样如此。首先从最基本的三个英文缩写开始。

AI 是英语 Artificial Intelligence 的缩写，中译为人工智能。

ML 是英语 Machine Learning 的缩写，中译为机器学习。

R.S. Michalski, J.G. Carbonell 与 T.M. Mitchell 所著《Machine Learning: An Artificial Intelligence Approach》的书名恰好可以作为二者之间关系的简单说明：机器学习是一种实现人工智能的途径。

DL 是 Deep Learning 的缩写，中译为深度学习，是一种机器学习技术，是机器学习的一个分支。

上述关系可以简单总结为人工智能包含机器学习及其他技术，如知识库(knowledge base)等。机器学习包含深度学习及其技术，如逻辑回归(logistic regression)、朴素贝叶斯(naïve Bayes)等。

Ian Goodfellow, Yoshua Bengio 与 Aaron Courville 所著《Deep Learning》中的第一章图1.4以维恩图的方式说明了上述关系。

发送 Venn, 维恩图 或 韦恩图 (三个关键词的任一个，不区分大小写)到微信公众号 **AI精研社**，可查看相关图示说明。

如果觉得上述略显枯燥，可以参考《人工智能极简编程入门》的第一章。作者团队在尝试用更风趣的方式为同学们建立起对AI、ML、DL的直观感受。

此外，周志华老师《机器学习》以1.5节、1.6节专业又易懂的语言进行了讲解，几乎不需要任何前置专业基础。

本小节貌似突然出现了满屏的英语，请不要因此把书扔掉，目前需要掌握的只有3个缩写，6个英文字母，同时也是本节的标题 AI、ML、DL，以及这3者之间的关系。

### 4 参考资料

Ref

上一节中提及了三本书，其中《机器学习》与《Deep Learning》是入门 ML 的主要参考资料。

由于本书及本方案频繁引用这些参考资料，为了方便讨论，采用以下缩写。

1. slm 表示 李航老师 所著《统计学习方法(第2版)》
2. mlb 表示 Tom M. Mitchell 所著《Machine Learning》
3. wml 表示 周志华老师 所著《机器学习》算法圈小伙伴亲切的称之为"西瓜书"
4. dlb 表示 [Ian Goodfellow, Yoshua Bengio 与 Aaron Courville](https://www.deeplearningbook.org/) 所著《Deep Learning》  
(<https://www.deeplearningbook.org/>) 算法圈小伙伴亲切的称之为"花书", "深度学习圣经"
5. mlapp 表示 Kevin Patrick Murphy 所著《Machine Learning: A Probabilistic Perspective》
6. prml 表示 Christopher M. Bishop 所著《Pattern Recognition and Machine Learning》
7. esl 表示 Trevor Hastie 所著《The Elements of Statistical Learning》
8. ngml 表示 吴恩达老师开设的 ML 系列课程，包含 coursera 上的慕课以及[斯坦福大学的 CS229](http://cs229.stanford.edu/notes/cs229-notes1.pdf)  
(<http://cs229.stanford.edu/notes/cs229-notes1.pdf>)
9. cmu601 表示 卡内基·梅隆大学开设的 10-601 ML 系列课程

具体引用如 wml.2，表示西瓜书第2页；slm sec 3.1 表示李航老师所著《统计学习方法(第2版)》3.1 节。

除此之外，以下未设计缩写的参考资料同样重要

李航老师的官网 (<http://www.hangli-hl.com/blogs.html>)

李航老师的新浪博客以及slm勘误表 (<http://blog.sina.com.cn/u/2060750830>)

wml勘误表 (<https://cs.nju.edu.cn/zhoush/zhoush.files/publication/MLbook2016.htm>)

谷歌发布的ML词汇表 (<https://developers.google.com/machine-learning/glossary>)

林轩田老师 机器学习基石 (<https://www.bilibili.com/video/av61680926/?p=2>)

发送 ref.ml 到微信公众号 **AI精研社**，可获取相关链接。

## 5 学习建议

### Advice

本书及本方案以亲自动手、生动体验为主要特点，因此同学们一定要亲自动手做笔记、推公式、撸代码。

笔记中需要标明是来自哪个参考资料，如 slm sec 3.1，如果不习惯使用缩写，可以直接注明，如"这条笔记来自李航老师所著《统计学习方法(第2版)》3.1节"。

如果是原文引用，可以通过段落格式注明，比较便捷的方法是使用markdown ">" 标记。

以下内容(不包含双引号)

"> k近邻算法简单、直观 ..."

在markdown环境下显示如下

k近邻算法简单、直观 ...

每条笔记都可以尝试，概念、公式、代码一一对应，这样不仅自己记忆深刻，也方便参考笔记的同学更容易更准确的理解。本书大量章节采用这种风格，为同学抛砖引玉。

在参考他人的笔记、视频、代码或其他形式的分享时，如果觉得好请随手点赞、转发，每一份支持都会成为分享者的动力，进而成为社区发展的动力；如果觉得不好，请提出具体的建议，改进的方法，如第几页第几段出现了什么错误，第几分第几秒讲的不够详细；如果觉得分享者的内容简单，很可能是因为同学的水平已经不再是纯小白，快进、跳过即可；如果觉得分享内容听不懂，可能是需要一定前置基础或额外参考资料，向分享者留言询问即可；情绪发泄式的抨击嘲讽对社区发展、对自己进步都不会有任何积极的作用；表达不满的方式有多种，其中较为积极的一种是亲自分享更优质的内容，供大家学习交流。

上述建议可以简单的总结为礼貌、平等、专业的交流。

学习的过程中将不可避免的遇到各种各样的问题，良好的习惯是先谷歌，查参考资料，再提问。

如果是概念性的问题，建议注明不少3处的参考资料，如果是代码类的问题，建议通过 colab 或 mybinder 提供一现复现的问题环境，对colab 或 mybinder 不了解的同学可以参考《人工智能极简编程入门》。

以上建议是作者团队在基于github、stackexchange的技术交流中总结经验，可以让自己问题得到更快更好的回答，让自己的分享得到更多的反馈。这些环节、套路最终的目标只有一个，帮助同学们掌握、融通算法思想并落地为可交付可维护的系统。

## 6 优雅快捷的写公式-latex

Why latex?

上一小节提及了笔记、问题等学习、交流环节。当前ML社区中的论文，基于github、stackexchange的技术交流，以及wml勘误表 (<https://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/MLbook2016.htm>)都是以Latex或mathjax格式发布的，因此建议同学们在提问、笔记中尝试使用Latex或mathjax。

本书的一些符号约定如果不使用latex，则会严重影响输入效率。如 $D_{n_j}$ ，其中 $n_j$ 是 $D_{n_j}$ 的下标，而 $j$ 又是 $n_j$ 的下标。

对 latex 不熟悉的同学可以访问以下URL

<https://www.overleaf.com> (<https://www.overleaf.com>)

集教程、在线运行、分享互动于一体的latex平台，非常符合所倡导的极简体验学习套路。

## 7 环境搭建

Dev Env

本书使用 Python3.7 + jupyter notebook 作为主要开发环境。对 jupyter notebook 不熟悉的同学可以参考《人工智能极简编程入门》第二章。

有软件开发环境搭建经验的同学可直接访问下URL，下载 anaconda，本地运行，根据提示安装即可。

[anaconda下载页](https://www.anaconda.com/distribution) (<https://www.anaconda.com/distribution>)

## 8 第一章补充说明

Supplementary of ch01

根据作者团队的有限经验，根据使用场景不同，符号约定的设计可能需要同时兼顾分享者(Latex输入、手写、代码、口述、习惯)与听众/读者(易读、易懂、减少歧义、方便记忆)双方或多方。

首先是最广泛的习惯。 $i, j, k, l, m, n$  按照英文字母表中的顺序排列，因此，在指代数据集时会自然的表示为， $m$ 行 $n$ 列， $m$ 个样本 $n$ 个特征，在指代数据时，会自然的表示为第 $i$ 个样本，第 $j$ 个特征。

如，在课堂上， $X_{sub\ i}$  比  $X_i$  更清楚；印刷或电子版(如pdf)中，上标中包含括号 $x^{(j)}$ 与不包含 $x^j$ 相比，前者可以减少初学者的混淆。

汉语与英语的语言习惯亦对符号约定有影响。如，第 $i$ 个样本的第 $j$ 个属性，符合汉语习惯，而英语中则可能是，the  $i$ th attribute in the  $j$ th training example。

都是遵循了  $i, j, k$  的先后顺序，但是指代的概念可能恰好相反。

$x_{ij}$ (wml.2) 与 Python 代码最为接近，方便书写与编程实现，但是在口述指代属性时， $x_{sub\ i}$ (cmu601)则更便捷清楚。

如果采用上标  $x^{(i)}$  表示  $i^{th}$  example，则 slme2 中的大量图示都需要进行相应调整，可能会增加初学者的学习成本。

本书采用  $x_i^{(j)}$  指代第 $i$ 个样本的第 $j$ 个属性，以  $X^{(j)}$  指代整个数据集(或partition划分)的第 $j$ 个属性，最大程度的延续slm的符号约定。

同学们在参考cmu601与ngml时，需要对调 $i$ 、 $j$ 的含义。

In [ ]:

1