

前言

写作这本教材的初心，就是想为本校人工智能本科专业的第一门机器学习相关课程，提供一本合适的教材。所谓“合适”主要体现在以下几个方面：

第一，理论和实践并重。既有理论的系统讲解、公式的详细推导，也有 Python 代码实现的详细讲解。以“2.3 对率回归”为例，我们从最简单、最基本的线性模型讲起，自然引入“含参模型”和“非参模型”的概念，进而针对分类问题导出核心概念——线性分类模型。接下来，详细推导对数几率和对数似然函数，以及负对数似然函数的最小化方法——梯度下降。有了对理论的透彻理解和对公式的详细推导，学生已经跃跃欲试、想要动手实践了。所以，紧接着就是代码的讲解，特别是核心代码的详细讲解。代码讲解过程中，特别强调与理论和公式的对应，一方面是为了进一步加深对理论和公式的理解，一方面也体现理论如何落地（比如如何把公式对应到 Numpy 库函数）以及编码过程中需要考虑的各种因素（比如计算效率）。

第二，模型和算法采用 Python 从 0 实现。不依赖于已有的机器学习库，从 0 开始实现，目的是把基本功练扎实。本教材只依赖于 Python、Numpy 和 Matplotlib，不依赖于已有的机器学习库，是学习机器学习的最小知识集。采用最小知识集有两个显著的优点：一是，学生的精力专注于核心知识，减少学习负担；二是，从 0 实现机器学习算法，更有利于把基本功练扎实。

第三，面向应用的习题和实验。以培养高水平应用型人才为导向，把丰富的应用融合到课程材料(特别是习题和实验)中，切实加强学生解决实际问题能力的培养。应用实例由浅入深、丰富而有趣。一些应用反复出现，有利于不同模型之间的比较和选择。

第四，强化算法思想讲解。算法的核心是思想，通过打比方、可视化等方式讲解算法思想及其发展过程，更有利于学生对其的深入理解和把握，从而进一步对其进行创造性的应用。还是以“2.3 对率回归”为例。讲解“梯度下降”这个经典优化算法的核心思想时，我们以“下山”来打比方——如何最快到达山脚，并通过可视化的方式说明“步长”的重要性。在此基础上，进一步拓展，比如一定能确保下到山脚吗？学生通过形象的方式理解了该算法的核心思想后，自然就会进一步思考和拓展，活学活用。

第五，突出独立思考、提出问题能力的培养。讲解过程中突出问题驱动下的思考过程，为什么提出这个问题/如何解决这个问题/又会产生什么新的问题。在习题和实验中，也会类似的进行引导，以培养学生独立思考、提出问题的能力。同样以“2.3 对率回归”为例。我们在讲解“含参模型”和“非参模型”的过程中，会提出“kNN 模型不是有个参数 k （近邻数）吗？”这个问题，引发学生的思考。在这个问题的驱动下，自然引入“超参数”这个概念。这其实就是问题驱动下的思考过程，目的是引导学生独立思考、提出问题。

教材内容简介

本教材共分为 9 章，涵盖了机器学习的基本内容，要求读者已经学习过 Python 程序设计基础，具备高等数学、线性代数、概率与统计相关数学知识。

第 1 章是概述。从整体上对“机器学习”进行一个概要性的介绍，从机器学习的基本概念开始，以机器学习的历史与现状收尾，谈及其术语、任务、分类等重要部分及其相互间的内在联系。虽然比较粗略，然期望达到“知其概貌”就足矣。

第 2 章是离散变量与分类模型。包括经典的 k -最近邻、决策树、对率回归、支持向量机，以及近 10 多年以来发展最为迅猛、应用最为广泛、影响最为深远的神经网络模型。

第 3 章是连续变量与线性回归模型。包括基本线性回归、岭回归、局部加权线性回归、

LASSO 回归。

第 4 章是维数灾难与降维。包括基本概念、主成分分析（给出最大化投影方差、最小化投影误差两种等价的推导方式）、奇异值分解。

第 5 章是 k 均值聚类。包括聚类的基本概念、k 均值聚类、k 均值++聚类。

第 6 章是生成模型与贝叶斯分类器。包括贝叶斯最优分类器、朴素贝叶斯分类器、半朴素贝叶斯分类器和贝叶斯网。

第 7 章是自监督与大语言模型。包括 Transformer、自监督学习、GPT 与大语言模型的预训练。

第 8 章是环境监督与强化学习。包括 ChatGPT 的三阶段训练流程、强化学习的形式化、策略最优化算法、环境构建与训练奖励模型。

第 9 章是实验部分。

关于学时分配

本校人工智能本科专业的《机器学习》课程共计 48 学时（28 学时理论+20 学时实验），各章学时数分配如下：

第 1 章 2 学时理论。

第 2 章 2.1 节~2.5 节，每节 2 学时理论+2 学时实验。

第 3 章 3 学时理论+2 学时实验。

第 4 章 4 学时理论+3 学时实验。

第 5 章 3 学时理论+3 学时实验。

第 6 章 2 学时理论+2 学时实验。

第 7 章 2 学时理论。

第 8 章 2 学时理论。

建议教师读者可根据自己课程的实际情况，参考上面的学时分配方案进行合理分配。

配套网站

本教材的配套网站为 <https://git.intcplab.cn/yuhuapeng/Machine-Learning-in-Python>，将在此处发布教材相关信息，包括实验用到的数据、部分习题解答、教材样章、教材勘误等。

致谢

首先，要特别感谢成都大学教务处、成都大学计算机学院、成都大学计算机学院人工智能课程协同体，对本教材从谋划、立项、再到出版全过程的大力支持！

其次，要特别感谢清华大学出版社计算机与信息分社，感谢社长魏江江和编辑葛鹏程两位老师的专业指导和辛苦工作！

最后，要感谢所有参与到本书从谋划、立项、再到出版全过程的各位老师、同学和朋友。

限于编者专业水平，错误和不当之处在所难免，烦请读者朋友们不吝赐教，定当虚心接受，并不胜感激！

编委

2024 年 3 月于成都