

高职院校《机器学习算法》课程基于 JupyterLab 的教学探索

黄伟纪 许君

DOI:10.12238/acair.v3i1.11917

[摘要]《机器学习算法》课程是人工智能技术应用专业的专业核心课程之一,该课程对线性代数、概率论与数理统计、微积分等高等数学知识的要求比较高,JupyterLab作为一种基于Web的交互式开发环境,为机器学习提供了丰富的第三方计算类库和友好的用户界面。本文根据高职院校学生高等数据知识掌握比较薄弱的特点,通过采用基于JupyterLab的教学模式,探索如何提升学生学习该课程的兴趣和学习效果,并在实际教学中取得了较好的效果。

[关键词] 机器学习算法; JupyterLab; 教学探索; Markdown

中图分类号: G95 **文献标识码:** A

The course "Machine Learning Algorithms" in vocational colleges is based on the teaching exploration of JupyterLab

Weiji Huang Jun Xu

[Abstract] The course of "Machine Learning Algorithm" is one of the core courses of artificial intelligence technology application, which has high requirements for advanced mathematical knowledge such as linear algebra, probability theory and mathematical statistics, calculus, etc., JupyterLab, as a Web-based interactive development environment, provides a rich third-party computing library and a friendly user interface for machine learning. By adopting the teaching mode based on JupyterLab, how to improve students' interest and learning effect in learning the course was explored, and good results were achieved in actual teaching.

[Key words] Machine learning algorithm; JupyterLab; instructional exploration; Markdown

引言

目前,高职院校《机器学习算法》课程的教学实践仍然处于摸索阶段,课程由于其自身难度以及高职院校学生高等数学基础知识较差的特点对于人工智能技术应用专业的课程建设提出了较大的挑战,《机器学习算法》课程属于多领域交叉学科,涉及线性代数、概率论与数理统计、微积分、算法理论、计算机编程等交叉知识点。如果没有掌握先导相关数学知识,学生既看不懂公式、无法推导公式,也难以将枯燥的公式和实际问题对应起来。高职院校学生的基本学情是学生高等数学基础较差、许多高职院校在上该课程前相关的高等数学课程开课不多。基于以上原因,高职院校《机器学习算法》在教学实践中面临较大的挑战。

1 基于JupyterLab的教学探索

1.1 JupyterLab简介

JupyterLab是一个集文本编辑器、终端以及各种个性化组件于一体的全能IDE,提供一个全面、灵活且高效的交互式计算环境,允许用户在一个统一的界面中组织多个文档和窗口,包括Python源代码、文本文件、图像、HTML等,并进行编辑、运行和协作。JupyterLab的特点包括:

(1)集成了丰富的第三方科学计算及机器学习类库,方便学

生在不需要掌握高等数学理论的基础上使用各类机器学习算法;(2)实时交互模式,可以直接输入代码并立即得到运行结果,同时运行结果嵌入到文档中,非常适合调试Python代码。(3)多种数据格式支持:可以查看并处理多种数据格式,进行丰富的可视化输出或Markdown形式输出。

1.2 基于JupyterLab的教学模式

1.2.1 JupyterLab集成了机器学习所需的第三方丰富的计算类库。JupyterLab工具内置集成了Numpy、Pandas、Matplotlib、Seaborn、Sklearn等机器学习所需的第三方类库,方便学生使用时不用自己下载安装。Numpy是Python用来做矩阵、高维数组运算的数学计算库;Pandas是Python用来做数据预处理、数据操作和数据分析的库;Matplotlib是数据可视化基础库;Seaborn是基于Matplotlib开发的数据统计可视化库;Sklearn库是机器学习库,它提供了从数据预测处理到模型训练的机器学习流程的各个步骤的实现方法,包括:数据分类、回归和聚类的机器学习算法,如:支持向量机、随机森林、K值聚类、Lasso回归、人工神经网络算法等。通过学习JupyterLab工具内置集成的以上五个机器学习类库的使用,使得同学们避免理解机器学习算法的繁琐数学推导过程,在没有线性代数、微积分等相关高等数学知识的背景

下专注各类机器学习算法的具体应用场景和使用方法。

1.2.2 利用Markdown标记语言在代码模块中内嵌帮助说明。JupyterLab编辑器每个编辑单元支持两种编辑模式：(1)采用Markdown标记语言编写格式化帮助文档；(2)采用Code编码模式编写Python代码。在代码内容里嵌入采用Markdown标记语言编写有格式的帮助说明，利用Markdown标记语言的特点，帮助说明目录层次清晰，格式美观。例如在Numpy库学习时在代码中增加Markdown标记语言编写的目录层次如下图(见图1)，每个目录节点展开就是其具体的使用案例，让学生在掌握学习Numpy库全貌的同时，逐步理解各个具体的功能实现方法。由面到点的学习步骤，让同学们更容易理解。

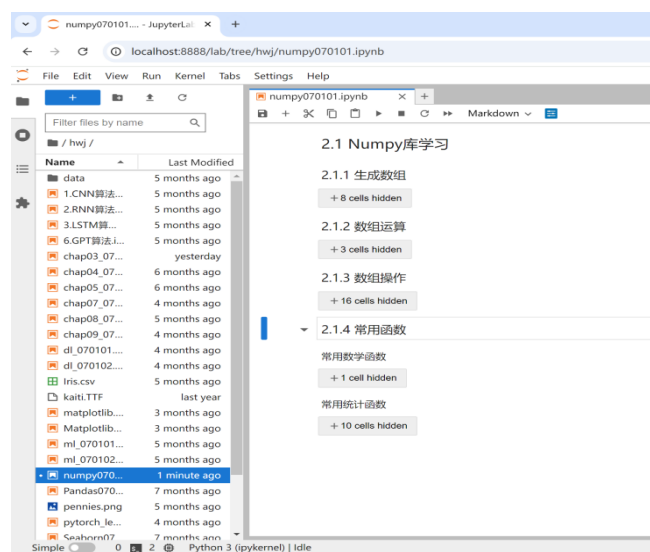


图1 Markdown编写的Numpy库学习帮助说明

1.2.3 代码逐行运行与数据可视化。JupyterLab工具提供代码实时交互运行模式，学生编写代码，可以逐行运行查看输出结果，同时运行结果内嵌保存在源代码文档中，方便老师批改同学的作业。通过逐步调试和优化算法，学生可以深入理解机器学习算法的工作步骤和工作原理。

JupyterLab工具利用Matplotlib和Seaborn库通过简单的几行代码就能方便进行数据可视化，学生通过图形可以直观地理解数据预处理和特征工程的重要性。以下是一个分组气泡图的展现及代码实现(见图2)，通过简单的四行代码就完成了分组气泡图的实现。

1.2.4 实训项目案例实践。《机器学习算法》课程最后阶段设计了实训项目案例：采用神经网络算法模型进行泰坦尼克号乘客存活情况预测。通过该项目案例的分析及实现加深学生对机器学习算法及机器学习处理流程的理解，同时还提高了他们的实际操作能力。学生利用JupyterLab平台，从数据收集开始，逐步完成数据预处理与探索、数据特征工程处理、模型训练、模型评估和模型优化等步骤。实训项目案例环节提高了学生的实践编程动手能力，也提高了学生学习的积极主动性，加深理解并掌握机器学习全流程处理过程(见图3)。

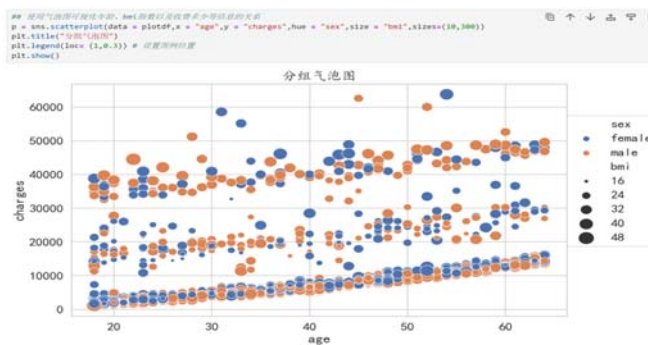


图2 分组气泡图的实现

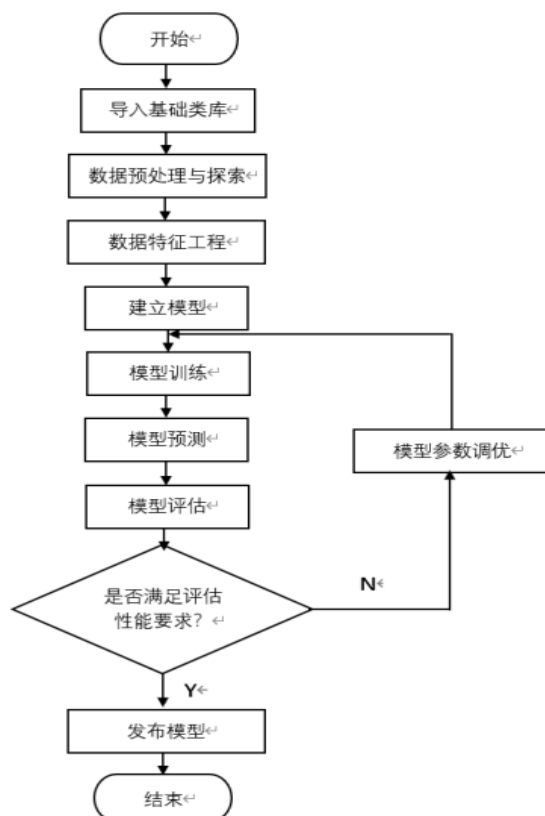


图3 机器学习处理流程图

以上的机器学习处理流程图采用Markdown标记语言编写有格式的帮助说明如下图(见图4)，通过Markdown帮助说明很好的展现了机器学习处理流程全貌，展开每个步骤节点就是其Python代码的具体实现，非常便于老师的讲解，同时学生也能非常直观的理解每个步骤的实现内容。

2 基于JupyterLab的教学实践效果

2.1 学生学习体验

通过使用JupyterLab平台，显著提升了学生在课程中的学习体验及学习兴趣。学生们在一个集成的环境中完成数据处理、算法实现和结果可视化，避免了繁琐的数学算法公式推导过程，通过JupyterLab平台内置的机器学习类库掌握了各类机器学习算法的使用和应用场景；JupyterLab的交互式运行模式使学生可以即时查看代码运行结果，有助于他们更好地理解和调试算法。

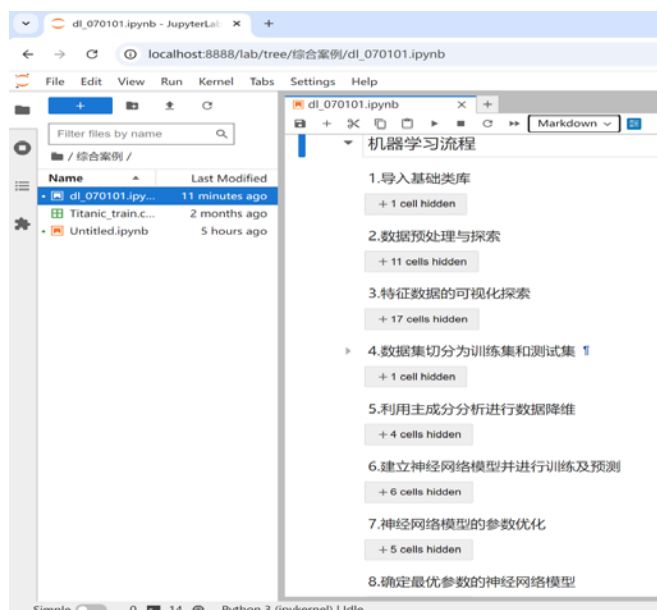


图4 Markdown编写的帮助说明

2. 2 学生学习成果

课程结束后, 通过对比学生的期末考试成绩和项目实践报告分析, 两个人工智能应用技术班的期末考试成绩分析如下图(见图5、图6), 人工1班的优良率为: 52. 3%, 人工2班的优良率为68. 2%, 不及格率均为0%, 课程成绩明显高于其他专业课的平均水平。学生们不仅掌握了经典的机器学习算法, 还能够独立完成从数据预处理到模型评估的完整项目流程。

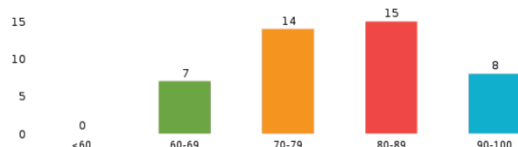


图5 人工智能技术1班期末考试成绩分布图

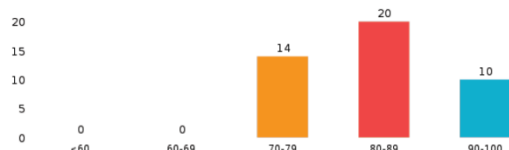


图6 人工智能技术2班期末考试成绩分布图

3 基于JupyterLab的教学展望

3. 1 增强交互式学习体验

利用JupyterLab的插件机制, 通过集成更多的交互式组件如: sql数据库连接查看组件、spreadsheet EXCEL表格组件、github

搜索github项目组件、Lsp代码补全和错误提示等组件, 丰富的插件显著更能提升使用JupyterLab工具的编码体验, 大大增强学生的学习效果和学习兴趣, 学生可以更加直观地理解复杂的编程和各类机器学习算法。使得基于JupyterLab的教学方式将更加增强学生的交互式学习体验。

3. 2 远程使用JupyterLab工具

由于机器学习编程需要的较多的硬件资源(特别深度学习算法需要较多的CPU/GPU的资源), 可以通过购置一台CPU/GPU资源较大的服务器安装JupyterLab并配置JupyterLab的远程访问模式, 这样学生端本机不用安装JupyterLab工具, 可直接通过浏览器访问服务器端的JupyterLab工具并能充分使用服务器的硬件资源进行深度学习算法的学习与实践。

4 结论

本文通过探讨高职院校《机器学习算法》课程基于JupyterLab的教学模式的内容、实践效果和教学展望, 结果表明, 基于JupyterLab的教学模式在《机器学习算法》教学方法中是一种行之有效的教学模式, 显著提升了学生对人工智能相关知识的学习兴趣, 大幅降低了高职院校学生学习该课程的畏难情趣。

[参考文献]

- [1]孙玉林. Python机器学习基础、算法与实战[M]. 北京: 化学工业出版社, 2023: 23-36.
- [2]末名编程. JupyterLab操作文档[Z]. https://blog.csdn.net/qj_44275213/article/details/139971514, 2024.
- [3]曾园园. 新工科背景下机器学习课程教学创新实践探索[J]. 计算机教育, 2022(11): 78-81, 86.
- [4]兰小艳. 高职院校机器学习课程的探索与实践[J]. 太原城市职业技术学院学报, 2022(2): 153-155.
- [5]郭兆麟, 徐婧. 高职机器学习算法课程教学方法探索与实践[J]. 电脑迷, 2023(11): 11.
- [6]徐川. 机器学习课程教学科研深度融合的探索与实践[J]. 新教育时代电子杂志(学生版), 2021, (7): 212, 217.

作者简介:

黄伟纪(1970--), 男, 广州涉外经济职业技术学院信息工程学院(鼎利产业学院)专职教师, 计算机工程师。研究方向: 人工智能及大数据技术及相关教学研究。

许君(1986--), 男, 广州涉外经济职业技术学院信息工程学院(鼎利产业学院)项目主任, 研究方向: 教务运营与管理。