**《概率论与数理统计》在机器学习中的案例教学研究**

 夏汉民

 （湖南大学，湖南，长沙，410081）

**摘要：**概率论与数理统计是机器学习最重要的分析工具之一，为顺应教学改革潮流和CDIO理念，本文将机器学习场景融入到概率论与数理统计教学中，研究和设计出合适的案例，采用启发式教学，给出了教学方案和教学过程。通过教学实践，促进了学生对概率论与数理统计的概念和基本原理的了解和掌握，激发了学生的学习兴趣，培养了学生的创新意识和独立思考的能力，提高了教学质量。

**关键词：**概率论与数理统计；机器学习；案例教学

**Abstract:** Probability theory and mathematical statistics are one of the most important analysis tools of machine learning. In order to conform to the teaching reform trend and CDIO concept, this paper integrates machine learning scenarios into probability theory and mathematical statistics teaching, studies and designs appropriate cases, adopts heuristic teaching, and gives the teaching plan and teaching process.  Through the teaching practice, it promotes the students to understand and master the concepts and basic principles of probability theory and mathematical statistics, stimulates the students' interest in learning, cultivates the students' innovative consciousness and independent thinking ability, and improves the teaching quality.

Key words: Probability theory and mathematical statistics;  Machine learning;  Case teaching

1. **序言**

 概率论与数理统计是理工科各专业和经济管理有关专业学生的一门必修公共基础课，但现有的概率论与数理统计教材，大多重理论轻应用，学生在学习中存在一些问题，其中最主要的就是对核心概念理解不透彻，不知道如何应用，也不知道如何编程实现。例如有些学生对独立事件和不相关事件的核心概念理解不透彻，分不清二者的区别，在对具体问题进行分析时不知道选择哪一个。另外，有些学生没有编程基础，无法实现算法。

 当前，机器学习的应用越来越广泛，而机器学习中的核心算法大多构筑在统计思维之上，概率统计在机器学习中的地位不言而喻。因此，通过机器学习的应用场景，让学生牢固掌握概率统计的核心概念和思想方法，提高学生解决实际问题的能力，促进学生创新能力的发展具有重要的意义。同时也是顺应教学改革潮流和CDIO理念，将概率统计理论体系与应用紧密结合，有利于培养学生的创新意识和独立思考能力，为学生以后的学习打下坚实的基础。

1. **机器学习场景中的概率统计原理**

概率统计本质上是利用数据发现规律、推测未知，而“发现规律、推测未知”正是机器学习的目标。机器学习的理论研究就是运用概率统计的基本原理和方法，分析学习任务的困难本质，为学习算法提供理论保证，并根据分析结果指导算法设计。其理论基础是PAC学习框架理论，我们只有了解和掌握了这个理论，才能更好地理解机器学习场景，选择和设计出好的案例，学生也才能更好在机器学习场景中理解概率统计的概念和基本原理，领悟运用所学知识解决实际问题的方法，激发学生的创新意识。

机器学习所研究的主要内容是关于在计算机上从数据产生“模型”的算法，即“学习算法”，有了学习算法，我们把经验数据提供给它，它就能基于这些数据产生模型，可以说机器学习是研究关于“学习算法”的学问。那么，机器学习首先就要考虑什么问题是可以高效学习的，哪些问题是难以学习的，完成学习需要多少实例以及能否找到一个通用的学习模型。这些都是机器学习的基本理论问题，也就是PAC学习框架理论必须解决的问题。只有深刻理解了这些理论，才能设计出符合实际问题的比较好的算法。该理论主要是采用概率统计原理和方法对学习算法进行研究，只有对相关的概率统计原理有一定的了解，才能掌握PAC学习框架理论，为学习算法的设计打下基础。因此，在学习概率统计时，将机器学习场景引入概

作者简介：夏汉民（1973-），男，湖南桃江人，湖南大学数学院讲师，研究方向：人工智能

基金项目：2021年湖南省普通高等学校教学改革研究一般项目“‘ 基于机器学习的概率论与数理统计案例教学研究’”（HNJG-2021-0341）

率统计实例教学，能促进学生对概率统计的基本原理和方法的掌握。

为了方便教学，首先要让学生充分了解机器学习场景，也就是要理解机器学习场景中要解决的一些基本问题，才能更好地将机器学习场景融入到概率统计的学习中。为此，我们认真学习了PAC学习框架理论，总结出了概率统计与机器学习之间的联系，如图一所示，从中可以看出概率统计在机器学习中的地位显而易见，是机器学习最重要的分析工具之一，从对机器场景的建模到算法设计和分析都可以运用概率统计的基本原理和方法。比如要快速准确对机器场景进行建模，就必须对样本的分别非常清晰；要估计某个统计参数，就必须掌握极大似然估计以及有偏性和无偏性；贝叶斯分类就必须知道贝叶斯公式和损失函数；在对算法的性能进行评估时，方差和期望要用得上。PAC学习框架理论告诉我们如何将机器学习场景融入到概率论与数理统计的课程教学中，让学生加深对这些知识的理解，为进一步学习打下基础。

在教学中，我们首先要认真分析机器场景中所用到的概率统计原理和方法，然后选择合适的教学案例，让学生在实际问题中如何应用这些原理和方法解决实际问题，促进学生对基本概念和原理的掌握。

图一 基于PAC框架理论教学方案

机器学习

基本概念

概率论与数理统计

概率，样本，样本分布 统计

概率统计原理，方差、期望的概念及应用

算法性能评估

泛化误差、经验误差

NFL定理

案例分析

算法

1. **案例教学过程**

案例教学主要以机器学习场景为背景，让学生能更好地掌握概率统计的基本概念，运用相关的基本原理和方法解决实际问题，培养学生的创新能力。教学过程仍以概率统计为侧重点，首先根据某个概率统计的原理和方法选择相应的机器场景，然后让学生了解该概率统计的原理和方法，并将其运用到机器学习场景中，对机器学习场景中的问题、算法和性能进行分析，最后再回到该概率统计原理和方法。通过这一过程让学生掌握概率统计的概念、原理和方法。教学过程如图二所示

编程验证

获取数据

选择机器学习案例

给出案例中的概率统计原理和方法，提出问题

分析论证

算法验证

思考总结，有针对性地提出问题

解决方案

图二 案例教学过程

在这个教学过程中，我们主要是采用启发式教学，将概率统计原理和方法融入到机器学习场景中，通过对一定的模拟数据和真实数据进行分析例证来得到相关的结论，引导学生掌握概率统计的基本概念，加深学生对所学知识的理解和运用。

概率统计和机器学习都是建立在大量数据的分析、整理和推断的基础上的，那么数据的获取是非常关键的。因此，在教学中我们必需介绍生成样本的常用方法，一种方法是从存储数据网站直接下载得到数据，比如UCI和Awesome公共数据集，UCI是机器学习的算法数据库，为数据集提供实时的更新维护，是科研工作者、教师和学生的机器学习的主要数据源。Awesome公共数据集是一个按照主题分类的，由社区公开维护的一系列数据集清单，比如生物学、经济学、教育学等。这里列出的大多数数据集都是免费的。还可以采用爬虫技术直接从网上抓取实时数据。另一种方法是通过计算机模拟生产，这就需要学生至少掌握一门编程语言，比如R语言或Python语言。还有一些常用的数据分析工具也要去了解，比如excel和spss。通过熟练运用这些工具，我们可以对获取的数据进行简单的分析，作为进一步处理和分析的基础。当然，为了进一步验证理论知识或者某个推断，我们要进行算法设计和编程，Python有强大的算法库，可以帮助学生设计实现各种算法，验证他们所学的知识和推断，激发学生的学习兴趣，有兴趣的学生还可以去多研究机器学习，促进对概率论与数理统计的基本原理和概念的了解和掌握。

通过基于机器学习场景的案例教学，学生对相关的概率论和数量统计的知识有了一定的了解后，我们还要根据所设计的场景以及学生对知识的掌握情况进行指导，提出新的问题，培养学生的创新能力。

1. **结语**

案例教学是概率论与数理统计很重要的教学环节，我们通过机器学习场景，把概率论与数理统计的基本原理和方法融入到机器学习案例中，让学生掌握好基本的概念和知识，培养了学生的学习兴趣和运用知识解决问题的能力。通过差不多一年的教学实践，学生对概念的理解和基本知识的掌握程度有了明显提高，在案例学习中学生能够领悟到将所学知识运用到实际问题中的那种美，激发了学生的学习积极性，有些学生还自主地进行机器学习算法的训练，积极参加数学建模竞赛，达到了较好的教学效果。

案例教学的关键是选择案例以及如何将概率论与数理统计的基本原理和方法与案例更好地融合，我们还在继续搜集案例，探索更好地融合。为了培养学生的创新意识和独立思考的能力，我们也将考虑给定某个相关的案例主题，让学生自己寻找和设计案例，经过分析思考找到问题的解决方案，并将这作为平时成绩的一部分，引导学生将所学知识运用到解决实际问题中，促进学生对概率论与数理统计的概念和基本知识的理解和掌握，同时也激发学生对机器学习的兴趣。

目前，基于机器学习的概率论与数理统计案例教学设计已有相当的进展，教学案例的选取和编写设计也有一定的成果。但要更好地将概率论与数理统计的基本原理和方法融入到机器学习场景中，还必须继续开展概率论与数理统计案例选取与设计方法的研究，通过基于机器学习的案例教学，引导学生将所学知识自觉地运用到实际问题中，促进学生对基本概念的理解，提高教学质量。

**参考文献：**

1. 张雨萌. 机器学习中的概率统计[M]. 北京：机械工业出版社，2021
2. 周志华. 机器学习[M]. 北京：清华大学出版社，2016
3. 金明. 概率论与数理统计实用案例分析[M]. 第二版，北京：中国统计出版社，2015
4. 王化琨. “数理统计”与“机器学习”的融合式教学的探索与实践[J]. 黑龙江教育（理论与实践），2022，2（1380）：65-66
5. 姜洋，衡红军，李俊生. 基于计算思维层次化认知的大学计算机教学改革实践[J]. 中国大学教学，2020（11）：59-63

作者简介：夏汉民（1973-），男，湖南桃江人，湖南大学数学院讲师，研究方向：人工智能

基金项目：2021年湖南省普通高等学校教学改革研究一般项目“‘ 基于机器学习的概率论与数理统计案例教学研究’”（HNJG-2021-0341）