

# Machine Learning

## 机器学习

In many ways, machine learning is the primary means by which data science manifests itself to the broader world. Machine learning is where these computational and algorithmic skills of data science meet the statistical thinking of data science, and the result is a collection of approaches to inference and data exploration that are not about effective theory so much as effective computation.

在很多情况下，机器学习是数据科学本身以及更广泛领域中的主要方法。机器学习是数据科学中计算及算法和统计思维相结合的产物，结果得到的是一整套的推理方法和数据分析工具。

The term "machine learning" is sometimes thrown around as if it is some kind of magic pill: *apply machine learning to your data, and all your problems will be solved!* As you might expect, the reality is rarely this simple. While these methods can be incredibly powerful, to be effective they must be approached with a firm grasp of the strengths and weaknesses of each method, as well as a grasp of general concepts such as bias and variance, overfitting and underfitting, and more.

术语“机器学习”有时候会被滥用就好像这是一剂灵丹妙药一样：*在你的数据上使用机器学习吧，你的所有问题都会得到解决。*虽然这些方法可能非常强大，但是你必须掌握每种方法的优缺点才能令它们更加有效，你需要掌握偏差和方差的基本概念，以及过拟合和欠拟合等等。

This chapter will dive into practical aspects of machine learning, primarily using Python's [Scikit-Learn](#) package. This is not meant to be a comprehensive introduction to the field of machine learning; that is a large subject and necessitates a more technical approach than we take here. Nor is it meant to be a comprehensive manual for the use of the Scikit-Learn package (for this, you can refer to the resources listed in [Further Machine Learning Resources](#)). Rather, the goals of this chapter are:

- To introduce the fundamental vocabulary and concepts of machine learning.
- To introduce the Scikit-Learn API and show some examples of its use.
- To take a deeper dive into the details of several of the most important machine learning approaches, and develop an intuition into how they work and when and where they are applicable.

本章会从实践的角度深入的介绍机器学习，主要使用Python的[Scikit-Learn](#)包。这并不是一个机器学习领域完整的介绍；因为这一课题过于庞大复杂，需要比本书更高级的技术书籍才能阐述清楚。本章甚至不是一个使用Scikit-Learn包的完整手册（为此，你需要在[更多机器学习资源](#)中列出的内容）。本章的目标是：

- 对机器学习基本术语和概念的介绍。
- Scikit-Learn包API的简单介绍以及使用例子。
- 对最重要的机器学习方法进行深入介绍，帮你建立它们工作原理的概念，并对它们的应用范围进行了解。

Much of this material is drawn from the Scikit-Learn tutorials and workshops I have given on several occasions at PyCon, SciPy, PyData, and other conferences. Any clarity in the following pages is likely due to the many workshop participants and co-instructors who have given me valuable feedback on this material over the years!

本章很多的材料都是从作者多次在PyCon、SciPy、PyData和其他论坛中对Scikit-Learn的教程和工坊中精选出来的。接下来的各小节内容都得到了这些论坛参与者和同事的反馈及帮助。

Finally, if you are seeking a more comprehensive or technical treatment of any of these subjects, I've listed several resources and references in [Further Machine Learning Resources](#).

最后，如果你在寻找更全面和深入的资料的话，在[更多机器学习资源](#)一节中列出了一些你需要的资源。