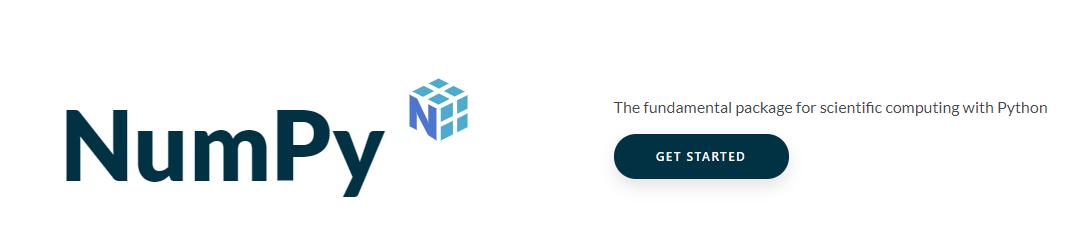
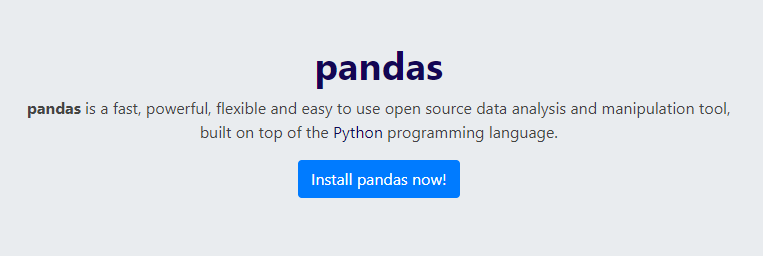
一 机器学习资源

1 机器学习常用三大Python库

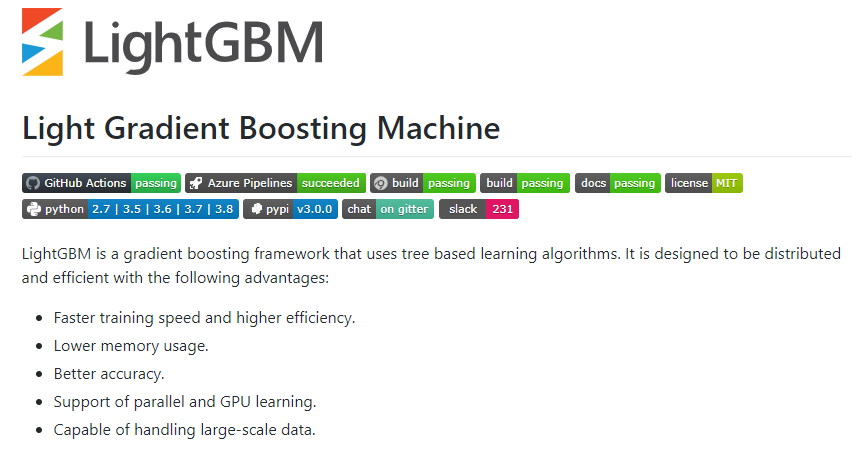


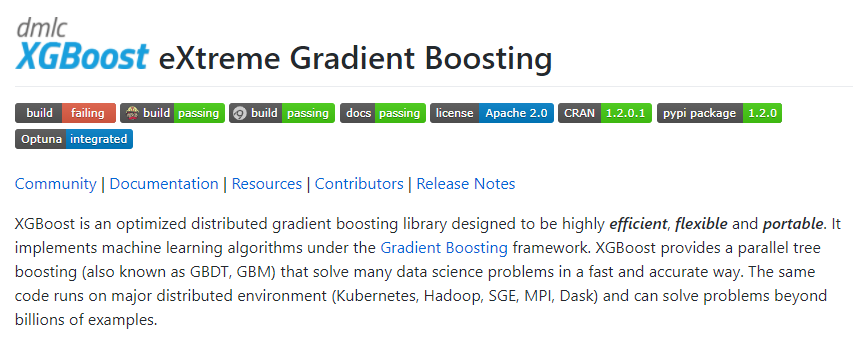




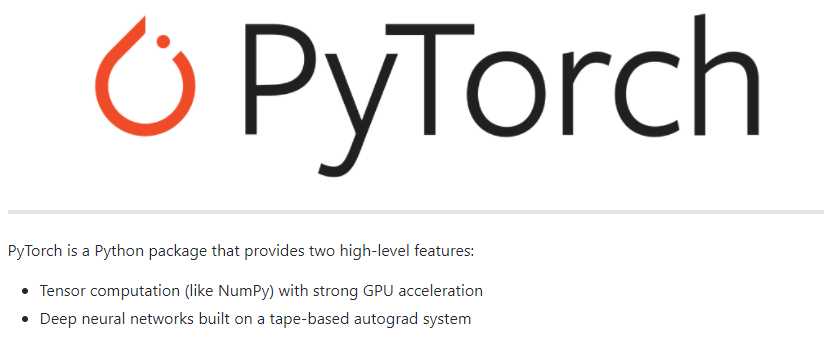
我们将以scikit-learn的功能作为主线，来学习机器学习的基本知识和工具。

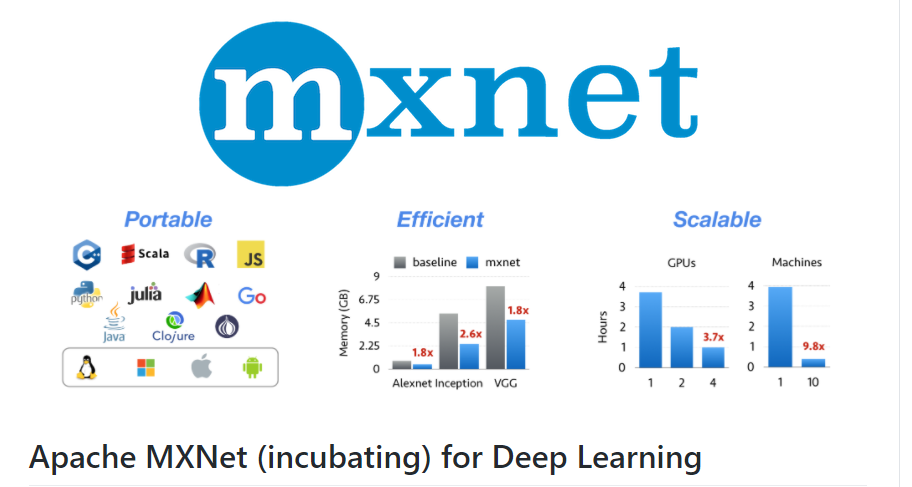
2 机器学习独立功能库



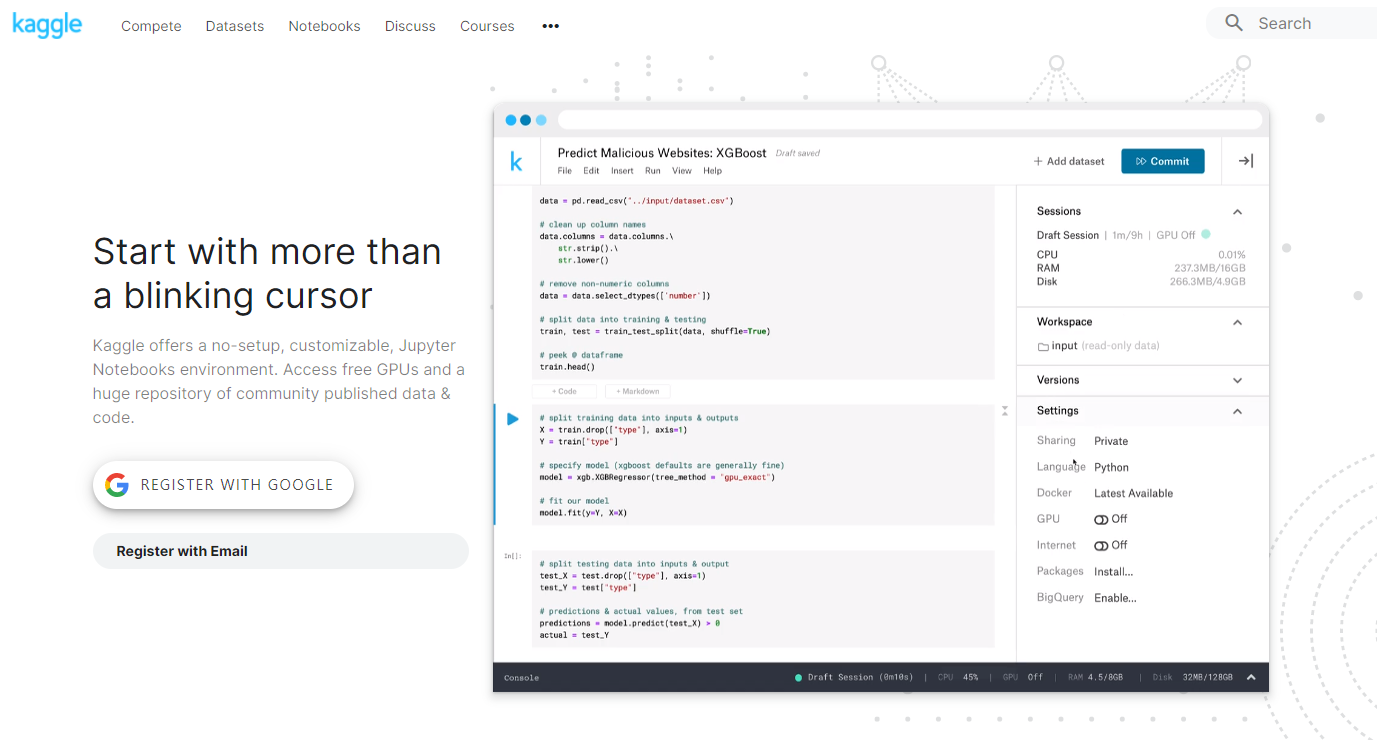


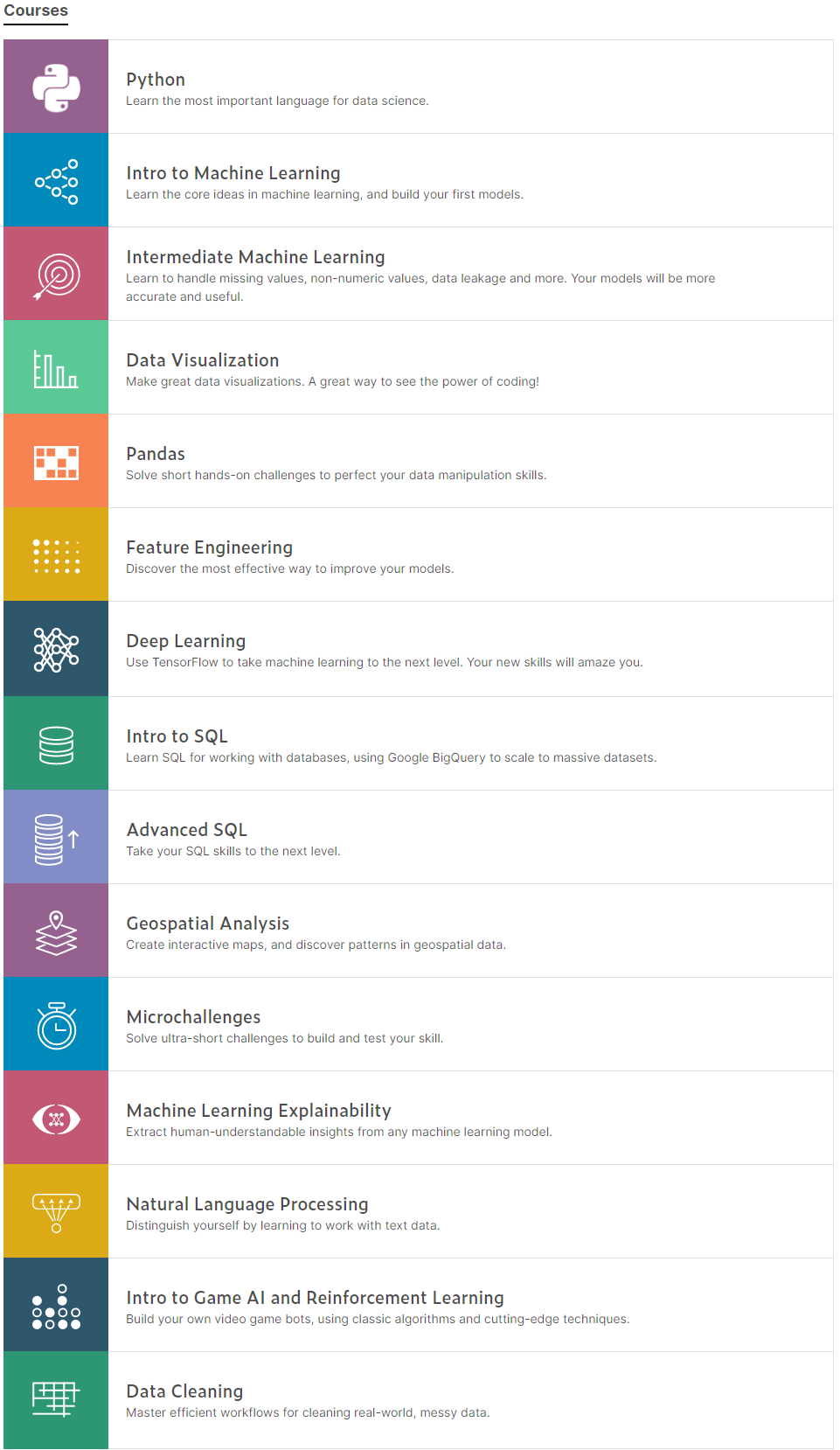




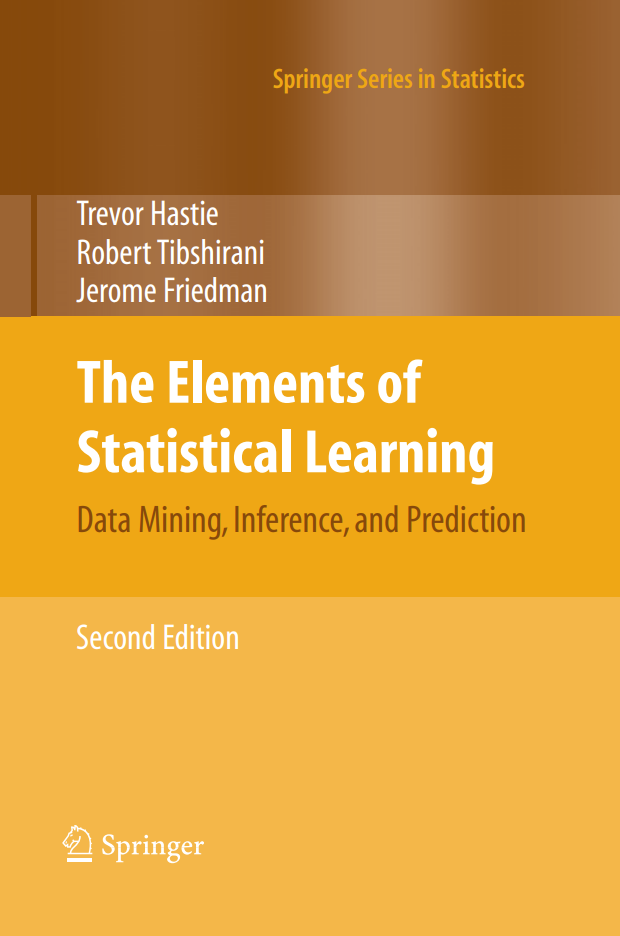


3 最适合学习的比赛网站





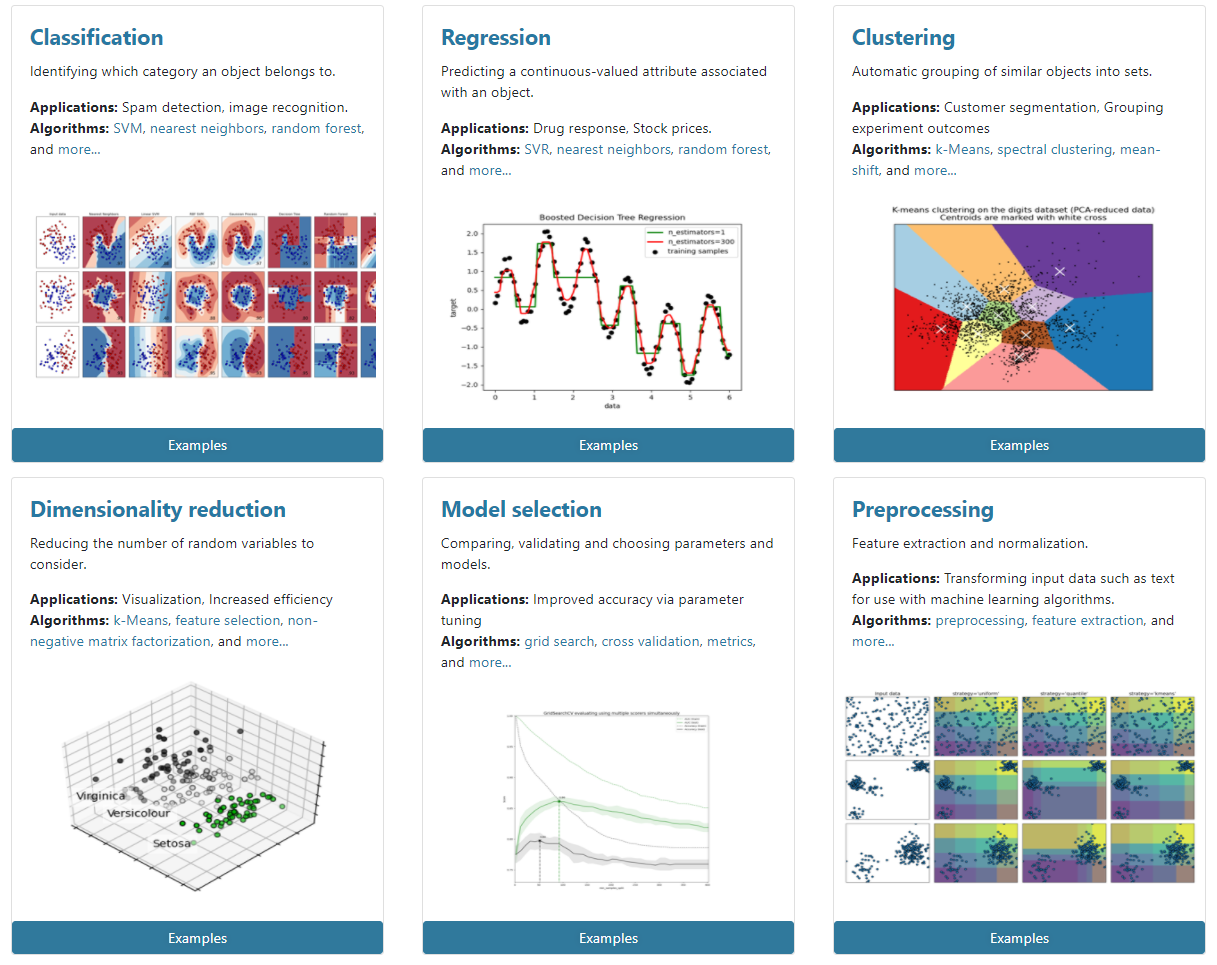
4 书籍



二 机器学习概念

机器学习，就是让机器从数据中学习知识，比如重要的模式和趋势，有意义的特征等。

这个定义非常笼统，其中包含的领域非常繁杂。下面是常常涉及的一些具体任务类型。



1 基本术语

X：输入，特征（feature），因子，自变量

Y：输出，标签（label），response，因变量

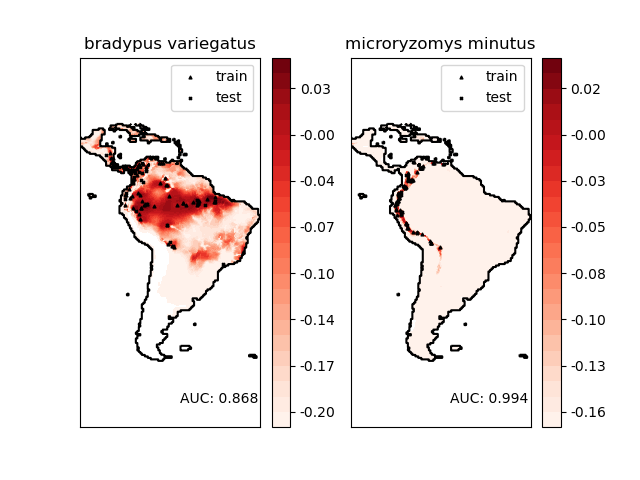
有Y的任务叫监督学习，没有Y的任务叫无监督学习。

Y是连续变量的一般叫拟合任务，Y是离散值通常是分类任务。

用于学习的数据叫训练数据（training data），不用于学习的数据叫测试数据或者校验数据（test data、 validation data）。

1 监督学习

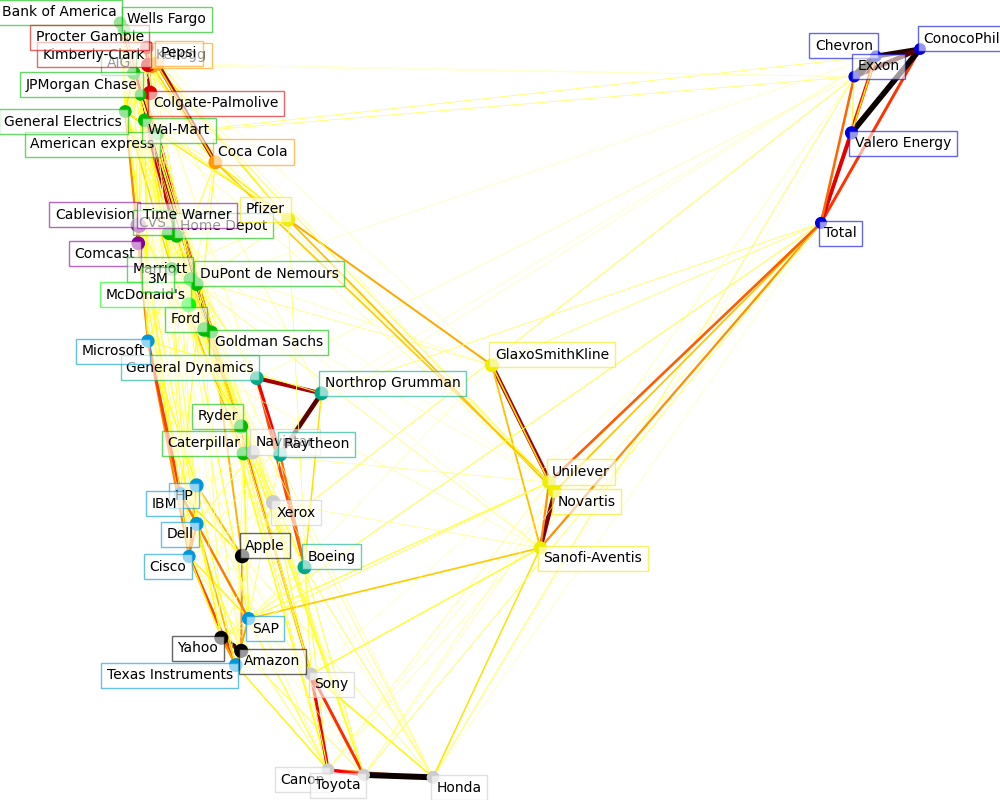
# Species distribution modeling



Example 1: Email Spam on

2 非监督学习

Visualizing the stock market structure



三 线性模型