

机器学习算法简介（准备知识）

机器学习算法入门

黄刚

1. Python

<https://docs.python.org/3/tutorial/index.html>

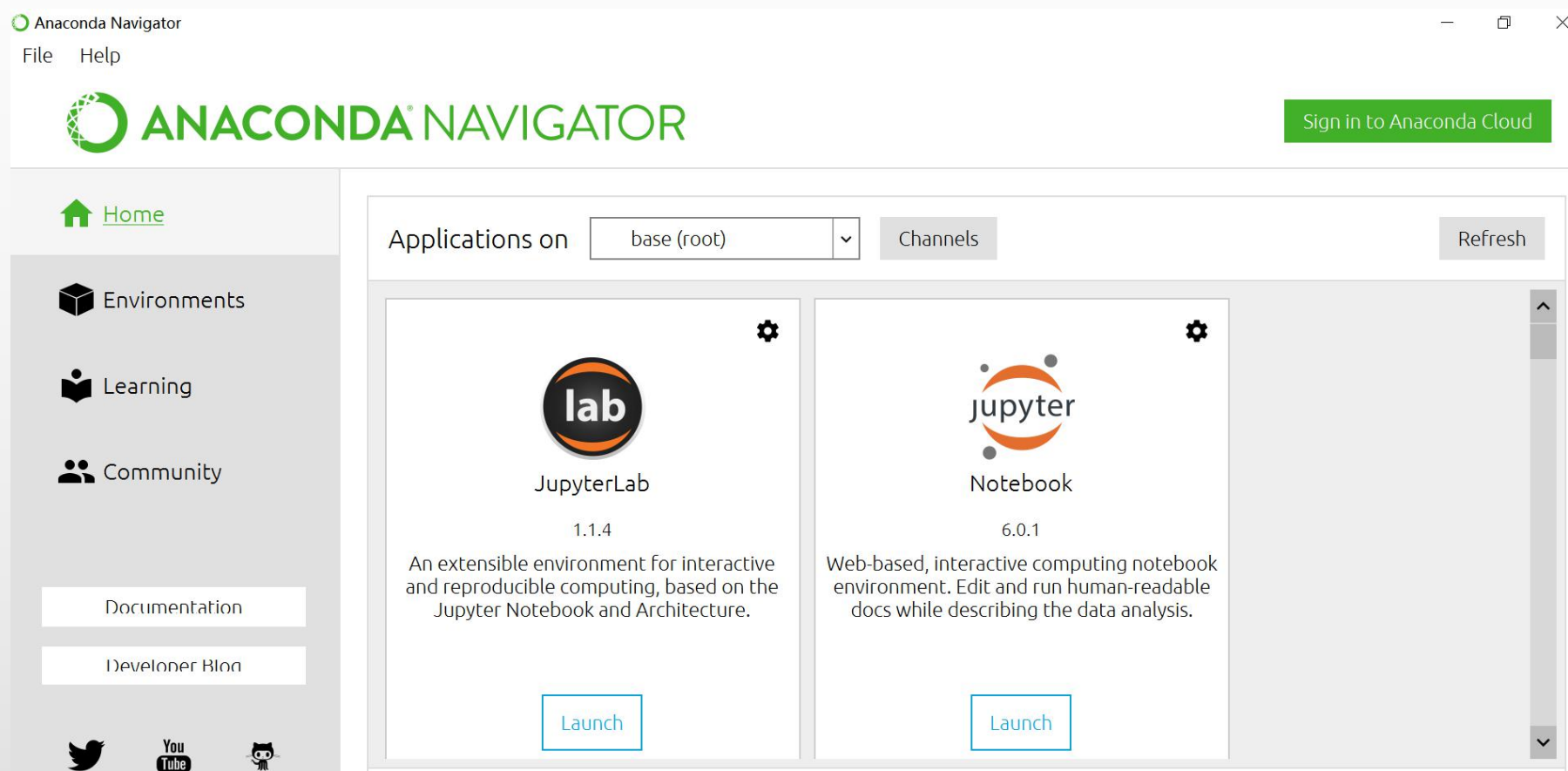
“Python is a programming language that lets you work quickly and integrate systems more effectively.” ----<https://www.python.org>

开发环境: Anaconda, Enthought, ...

用途: 科学计算, 数据分析, 网页编程, GUI开发, 软件开发,

Python环境：Anaconda

- Anaconda 使Python的环境创建与管理变得极其简单.



数据分析工具

- Numpy : <http://www.numpy.org/> (高维数组)
- Matplotlib: <https://matplotlib.org/> (数据分析作图)
- Pandas : <http://pandas.pydata.org/> (数据分析工具)
- Scipy: <https://www.scipy.org/> (数学、科学、工程)
- Scikit: <https://scikit-learn.org/> (机器学习)
- TensorFlow: <https://tensorflow.google.cn/> (机器学习)

2. 机器学习

- 机器学习 (Machine Learning) 是一门关于从数据中提取知识的学问 (统计学习). 它是当今AI的必然趋势.
- 机器学习是关于如何创建能从数据和观测中提高我们的经验的计算机程序 (Tom Riccio).
- 你想要能教计算机如何学习并如何提高经验. 这是机器学习之核心.

2.1 机器学习在科学上的应用

- 材料科学：研究材料的"结构--功能"对应关系
- 粒子物理：发现新的粒子
- 天文学：发现新的天体
- 化学：寻找最高效的化学反应,预测化学反应的产物以及预测晶体制备策略[1]
- 高分子物理化学：研究多尺度下高分子的结构和力学机制
- 生物信息学：基因序列分析（DNA序列编码特征分析，发现模式—功能对应关系等）

2.2 机器学习的类别(1): 非监督学习

- 机器学习可分为监督学习，非监督学习和强化学习等。
- 非监督学习 (unsupervised learning) 中，我们**没有**（或不必要）为实例做**标签**，我们的数据集中的实例是没有标签的。换句话说，在学习的过程中，模型 (agent) **得不到明确的反馈**，它的目的是从输入数据中学习某些模式 (patterns)。
- 例如，我们有顾客的数据，却没有任何类型的标签与之相联系。要解决的问题：给定这些数据点，我们能找出这些实例的数据点的聚类 (clusters) 吗？我们要寻找一个函数 F 以把输入集合 X 映射到聚类的集合 y 。这是非监督的算法。
- 有很多不同方法来实现非监督学习。最主要的有**聚类**，**降维算法**。

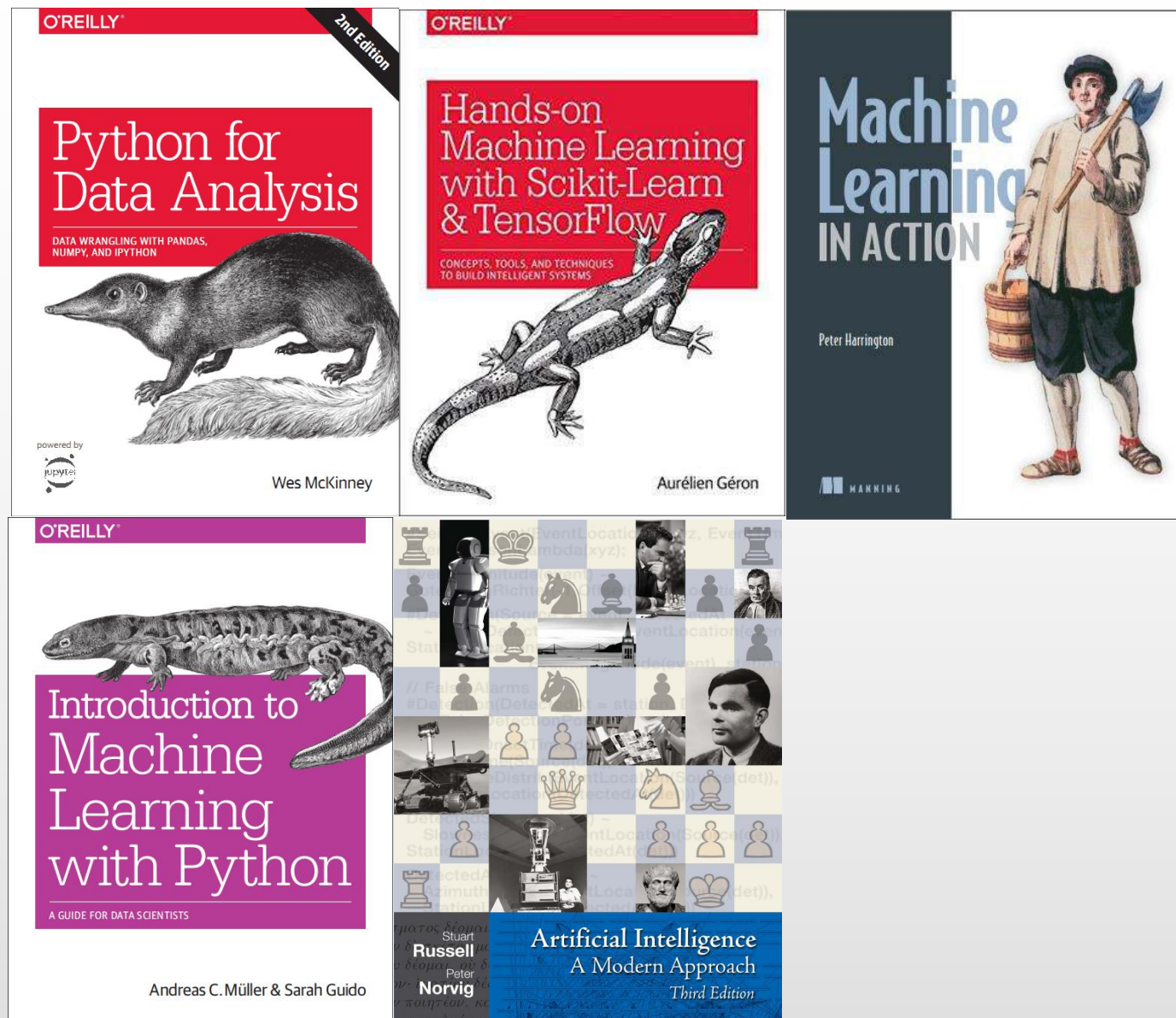
2.2 机器学习的类别(2): 强化学习

- 在强化学习 (Reinforcement Learning) 中, 我们设计智能体在随机或特定环境中演化.
- 智能体从**强化或延迟奖励 (惩罚)** 中学习 (从**环境**得到了一些**反馈**). 它是一类用于在输出结果为随机值的决策问题中的学习方法, 它包含了一个能连续计划、**学习并影响其环境**的智能体. 强化学习的驱动力是最大化奖励. 例如, 用于下棋的模型在一个回合结束后赢了1分, 它就知道它在前面的步骤中的某些操作做对了.
- 强化学习在**对抗, 博弈, 游戏, 投资, 交易**等等领域都可以应用. (举例)

2.2 机器学习的类别(3): 监督学习

- 监督学习 (supervised learning): 当你有**标签**时, 那么你做的就是监督学习. 此时, 智能体能得到**非常明确的反馈**. 这些**标签**可以是连续数值, 也可以是离散值. 我们想要建立的是一个函数, 给定一个输入集合或是实例的描述, 得出输出集合.
- 例如, 电子邮件分类, 人脸的识别, 手写体识别等.
- 如何找出分开这两类特征的**边界**是这类问题的目标. **监督学习**包括k近邻算法, 神经网络, 线性回归, 决策树算法等等.

3.参考资料



3. 参考资料

- [1] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk and M. B. Wenny, *Machine-learning-assisted materials discovery using failed experiments*. **Nature**, 2016 , **533** (7601) :73
DOI: 10.1038/nature17439
- [2] Luger, G. F. and Stubblefield, W. A., *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, Benjamin/Cummings, Menlo Park, CA, 1993.
- [3] A. C. Mueller, S. Guido, *Python机器学习入门*, O' Reilly, 2016
- [4] Russel, S. and Norvig, P., *Artificial Intelligence: A Modern Approach*, Prentice-Hall, Englewood Cliffs, NJ, 1995
- [5] Andrew Ng, Machine Learning (<https://www.bilibili.com>)