前言

Its goal is to give you the concepts, the intuition, and the tools you need to actually implement programs capable of learning from data.

本书目标是教会你从数据中学习的概念，直观感受，和一些你需要用到的工具。

本书代码库

<https://github.com/ageron/handson-ml>

本书要求：

懂得基本python语法

熟悉Numpy，Pandas，Matplotlib

数学要求：

calculus, linear algebra, probabilities, and statistics

微积分，线性代数，概率论，统计学

RaodMap

本书分为两部分：

**第一部分**

* Ml是什么？解决什么问题？ML系统的主要领域和关键概念是什么？
* 一个典型MP项目的主要步骤
* 通过调试模型来学习
* 优化损失函数
* 处理，清洗和准备数据
* 选择和设计（创建）特征
* 选择一个模型并通过交叉验证法调参
* ML的核心挑战，尤其是过拟合和欠拟合（偏差和方差折衷）
* 减少训练数据的维度，以避免维度灾难
* 最常见的算法:

Linear and Polynomial Regression, 线性和多项式回归

Logistic Regression,

k-Nearest Neighbors,

Support Vector Machines,

Decision Trees, Random Forests，

Resemble methods.

**第二部分**

神经网络和深度学习，包含以下主题

* 什么是神经网络？有什么好处？
* 使用tensorflow来建立和调试神经网络
* 最重要的神经网络架构：

Feedforward neural nets, 正反馈神经网络

Convolutional nets, 卷积神经网络

Recurrent nets, 周期网络

Long short-term memory (LSTM) nets 长短时记忆网络

Autoencoders 自编码器

其他学习资源推荐

Andrew Ng’s ML course on Coursera 吴恩达的课

Geoffrey Hinton’s course on neural networks and Deep Learning

ML博客文章：

<https://www.quora.com/What-are-the-best-regularly-updated-machine-learning-blogs-or-resources-available>

深度学习网站：

<http://deeplearning.net/>

<https://www.kaggle.com/>

**第一部分 The Fundamentals of Machine Learning**

**第一章.The Machine Learning Landscape 机器学习大观**

ML的定义：

ML is the science (and art) of programming computers so they can learn from data.

让机器能够从数据中学习。

机器学习类型

* 是否需要人工监督：

supervised（监督学习）,

unsupervised(非监督学习),

semisupervised（半监督）,

Reinforcement Learning(加强学习)）

* 是否能够联机递增（在线学习，批次学习）
* 是将新数据和已知数据作比较，还是根据训练数据建立模型（基于实例，还是基于模型）

**Supervised Learning 监督学习算法**

核心特征，有label。

分类问题经常用到。

常见的监督学习算法：

* k-Nearest Neighbors k-邻近算法
* Linear Regression 线性回归
* Logistic Regression 逻辑回归
* Support Vector Machines(SVMs) 支持向量机
* Decision Trees and Random Forests 决策树和随机森林
* Neural networks 神经网络

监督学习例子：

垃圾邮件拦截

**Unsupervised无监督学习算法**

**Clustering 聚类问题**

k-Means

Hierarchical Cluster Analysis (HCA)  层序聚类分析

Expectation Maximization 期望最大化

**Visualization and dimensionality reduction** **可视化和降维**

Principal Component Analysis (PCA) 主成分分析

Kernel PCA

Locally-Linear Embedding （LLE） 局部线性嵌入

t-distributed Stochastic Neighbor Embedding(t-SNE) t分布随机邻居嵌入

**Association rule learning** **关联规则学习**

**Apriori** 先验学习

无监督学习例子：

购物商场可视化处理数据

博客读者分析

常用见的3个无监督学习任务

Dimensionality reduction 降维也叫feature extraction：将相互有关系的几个特征合并到一个维度。

Anomaly detection 异常检测，比如检测到异常数据自动从数据集中删除。

Association rule learning 关联规则学习，主要目标是通过大量数据，发现其中的相关性。