

前言

这几年机器学习这种从经验学习的软件技术重现光明。在计算机诞生的早期，机器学习的概念已经出现，各种理论天马行空，限于计算成本而未能普及。随着计算设备的普及，日常生活中越来越多的机器学习应用，可以说它的成功开始变得习以为常。新应用如雨后春笋一般出现，很多都从机器学习中获得动力。

在这本书里，我们将看到一些机器学习的模型和算法。我们会介绍一些常用的机器学习任务和模型的效果评估方法。而这些模型和算法都是通过十分流行的Python机器学习库scikit-learn来完成，里面有许多机器学习的模型和算法，每个API都简单易用。

本书特点主要有：

- 内容通俗易懂。本书只需要基本的编程和数学知识
- 案例实用。本书的案例都很容易上手，读者可以调整后解决自己的问题。

本书内容简介

第1章，机器学习基础 (<http://muxuezi.github.io/posts/1-the-fundamentals-of-machine-learning.html>)，将机器学习定义成一种通过学习经验改善工作效果的程序研究与设计过程。其他章节都以这个定义为基础，后面每一章里介绍的机器学习模型都是按照这个思路解决任务，评估效果。

第2章，线性回归 (<http://muxuezi.github.io/posts/2-linear-regression.html>)，介绍线性回归模型，一种解释变量和模型参数与连续的响应变量相关的模型。本章介绍成本函数的定义，通过最小二乘法求解模型参数获得最优模型。

第3章，特征提取与处理 (<http://muxuezi.github.io/posts/3-feature-extraction-and-preprocessing.html>)，介绍了常见的机器学习对象如文本，图像与分类变量的特征提取与处理方法。

第4章，从线性回归到逻辑回归 (<http://muxuezi.github.io/posts/4-from-linear-regression-to-logistic-regression.html>)，介绍广义线性回归模型如何解决分类任务。将逻辑回归模型与特征提取技术结合起来实现一个垃圾短信分类器。

第5章，决策树——非线性回归与分类 (<http://muxuezi.github.io/posts/5-nonlinear-classification-and-regression-with-decision-trees.html>)，介绍了一种回归和分类的非线性模型——决策树。用决策树集成方法实现了一个网页广告图片屏蔽器。

第6章，K-Means聚类 (<http://muxuezi.github.io/posts/6-clustering-with-k-means.html>)，介绍非监督学习的K-Means聚类算法，并与逻辑回归组合起来实现一个照片分类器。

第7章，用PCA降维 (<http://muxuezi.github.io/posts/7-dimensionality-reduction-with-pca.html>)，介绍另一种非监督学习任务——降维。我们用主成分分析实现高维数据的可视化，建立一个脸部识别器。

第8章，感知器 (<http://muxuezi.github.io/posts/8-the-perceptron.html>)，介绍一种实时的，二元分类器——感知器。后面两章都是针对感知器的缺点发展起来的。

第9章，从感知器到支持向量机 (<http://muxuezi.github.io/posts/9-from-the-perceptron-to-support-vector-machines.html>)，介绍支持向量机，是一种有效的非线性回归与分类模型。我们用支持向量机识别街景照片中的字母。

第10章，从感知器到人工神经网络 (<http://muxuezi.github.io/posts/10-from-the-perceptron-to-artificial-neural-networks.html>)，介绍了人工神经网络，是一种强大的有效的非线性回归与分类模型。我们用人工神经网络识别手写数字。