

《机器学习实践：测试驱动的开发》

图书基本信息

书名：《机器学习实践：测试驱动的开发方法》

13位ISBN编号：9787115396183

出版时间：2015-8

作者：[美] Matthew Kirk

页数：204

译者：段 菲

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu111.com

《机器学习实践：测试驱动的开发》

内容概要

本书主要介绍如何将测试驱动开发运用于机器学习算法。每一章都通过示例介绍了机器学习技术能够解决的有关数据的具体问题，以及求解问题和处理数据的方法。具体涵盖了测试驱动的机器学习、机器学习概述、K近邻分类、朴素贝叶斯分类、隐马尔可夫模型、支持向量机、神经网络、聚类、核岭回归、模型改进与数据提取等内容。通过学习本书，你将能够利用机器学习技术解决涉及数据的现实问题。

《机器学习实践：测试驱动的开发》

作者简介

Matthew Kirk

是Modulus 7公司的创始人，该公司针对数据科学和Ruby开发提供咨询服务。Matthew从事程序设计工作已15年有余，在全球许多技术大会上做过机器学习和数据科学主题的演讲。

书籍目录

前言	xi
第1章 测试驱动的机器学习	1
1.1 TDD的历史	2
1.2 TDD与科学方法	2
1.2.1 TDD可构建有效的逻辑命题	3
1.2.2 TDD要求你将假设以文字或代码的形式记录下来	5
1.2.3 TDD和科学方法的闭环反馈机制	5
1.3 机器学习中的风险	5
1.3.1 数据的不稳定性	6
1.3.2 欠拟合	6
1.3.3 过拟合	7
1.3.4 未来的不可预测性	8
1.4 为降低风险应采用的测试	8
1.4.1 利用接缝测试减少数据中的不稳定因素	8
1.4.2 通过交叉验证检验拟合效果	9
1.4.3 通过测试训练速度降低过拟合风险	10
1.4.4 检测未来的精度和查全率漂移情况	11
1.5 小结	11
第2章 机器学习概述	13
2.1 什么是机器学习	13
2.1.1 有监督学习	13
2.1.2 无监督学习	14
2.1.3 强化学习	15
2.2 机器学习可完成的任务	15
2.3 本书采用的数学符号	16
2.4 小结	16
第3章 K近邻分类	17
3.1 K近邻分类的历史	18
3.2 基于邻居的居住幸福度	18
3.3 如何选择K	21
3.3.1 猜测K的值	21
3.3.2 选择K的启发式策略	21
3.3.3 K的选择算法	24
3.4 何谓“近”	24
3.4.1 Minkowski距离	25
3.4.2 Mahalanobis距离	26
3.5 各类别的确定	27
3.6 利用KNN算法和OpenCV实现胡须和眼镜的检测	29
3.6.1 类图	29
3.6.2 从原始图像到人脸图像	30
3.6.3 Face类	33
3.6.4 Neighborhood类	36
3.7 小结	43
第4章 朴素贝叶斯分类	45
4.1 利用贝叶斯定理找出欺诈性订单	45
4.1.1 条件概率	46
4.1.2 逆条件概率	47

4.2	朴素贝叶斯分类器	48
4.2.1	链式法则	48
4.2.2	贝叶斯推理中的朴素性	49
4.2.3	伪计数	50
4.3	垃圾邮件过滤器	51
4.3.1	类图	51
4.3.2	数据源	52
4.3.3	Email类	52
4.3.4	符号化与上下文	55
4.3.5	SpamTrainer类	56
4.3.6	通过交叉验证将错误率最小化	63
4.4	小结	66
第5章	隐马尔可夫模型	67
5.1	利用状态机跟踪用户行为	67
5.1.1	隐含状态的输出和观测	69
5.1.2	利用马尔可夫假设简化问题	70
5.1.3	利用马尔可夫链而非有限状态机	71
5.1.4	隐马尔可夫模型	71
5.2	评估：前向-后向算法	72
5.3	利用维特比算法求解解码问题	75
5.4	学习问题	76
5.5	利用布朗语料库进行词性标注	76
5.5.1	词性标注器的首要问题：CorpusParser	77
5.5.2	编写词性标注器	79
5.5.3	通过交叉验证获取模型的置信度	86
5.5.4	模型的改进方案	88
5.6	小结	88
第6章	支持向量机	89
6.1	求解忠诚度映射问题	89
6.2	SVM的推导过程	91
6.3	非线性数据	92
6.3.1	核技巧	92
6.3.2	软间隔	96
6.4	利用SVM进行情绪分析	97
6.4.1	类图	98
6.4.2	Corpus类	99
6.4.3	从语料库返回一个无重复元素的单词集	102
6.4.4	CorpusSet类	103
6.4.5	SentimentClassifier类	107
6.4.6	随时间提升结果	111
6.5	小结	111
第7章	神经网络	113
7.1	神经网络的历史	113
7.2	何为人工神经网络	114
7.2.1	输入层	115
7.2.2	隐含层	116
7.2.3	神经元	117
7.2.4	输出层	122
7.2.5	训练算法	122

7.3	构建神经网络	125
7.3.1	隐含层数目的选择	126
7.3.2	每层中神经元数目的选择	126
7.3.3	误差容限和最大epoch的选择	126
7.4	利用神经网络对语言分类	127
7.4.1	为语言编写接缝测试	129
7.4.2	网络类的交叉验证	132
7.4.3	神经网络的参数调校	135
7.4.4	收敛性测试	136
7.4.5	神经网络的精度和查全率	136
7.4.6	案例总结	136
7.5	小结	136
第8章	聚类	137
8.1	用户组	138
8.2	K均值聚类	139
8.2.1	K均值算法	139
8.2.2	K均值聚类的缺陷	140
8.3	EM聚类算法	141
8.4	不可能性定理	142
8.5	音乐归类	142
8.5.1	数据收集	143
8.5.2	用K均值聚类分析数据	144
8.5.3	EM聚类	146
8.5.4	爵士乐的EM聚类结果	149
8.6	小结	151
第9章	核岭回归	153
9.1	协同过滤	153
9.2	应用于协同过滤的线性回归	154
9.3	正则化技术与岭回归	157
9.4	核岭回归	158
9.5	理论总结	158
9.6	用协同过滤推荐啤酒风格	159
9.6.1	数据集	159
9.6.2	我们所需的工具	159
9.6.3	评论者	162
9.6.4	编写代码确定某人的偏好	164
9.6.5	利用用户偏好实现协同过滤	166
9.7	小结	167
第10章	模型改进与数据提取	169
10.1	维数灾难问题	169
10.2	特征选择	171
10.3	特征变换	173
10.4	主分量分析	175
10.5	独立分量分析	177
10.6	监测机器学习算法	179
10.6.1	精度与查全率：垃圾邮件过滤	179
10.6.2	混淆矩阵	181
10.7	均方误差	182
10.8	产品环境的复杂性	183

《机器学习实践：测试驱动的开发》

10.9	小结	183
第11章	结语	185
11.1	机器学习算法回顾	185
11.2	如何利用这些信息来求解问题	186
11.3	未来的学习路线	187
作者介绍		188
封面介绍		188

《机器学习实践：测试驱动的开发》

精彩短评

- 1、Github 上的 Ruby 代码木有 README 完全不知道怎么跑起来...
- 2、Ruby 的测试代码没看太懂，但给我传达了测试的思想

《机器学习实践：测试驱动的开发》

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:www.tushu111.com