

精通机器学习



[精通机器学习_下载链接1](#)

著者:[美] Cory Lesmeister

出版者:人民邮电出版社

出版时间:2018-3

装帧:平装

isbn:9787115477781

机器学习是近年来的热门技术话题，R语言是处理其中大量数据的有力工具。本书为读者提供机器学习和R语言的坚实算法基础和业务基础，内容包括机器学习基本概念、线性回归、逻辑回归和判别分析、线性模型的高级选择特性、K最近邻和支持向量机等，力图平衡实践中的技术和理论两方面。

作者介绍:

Cory Leismester

具有十多年量化管理经验，目前在银行业担任高级量化管理经理，负责构建市场营销和监管模型。曾在礼来公司就职16年，负责销售、市场调查、精益六西格玛、营销分析、新产品预测等工作。

目录: 第1章 成功之路 1

1.1 流程 1

1.2 业务理解 2

1.2.1 确定业务目标 3

1.2.2 现状评估 4

1.2.3 确定分析目标 4

1.2.4 建立项目计划 4

1.3 数据理解 4

1.4 数据准备 5

1.5 建模 5

1.6 评价 6

1.7 部署 6

1.8 算法流程图 7

1.9 小结 10

第2章 线性回归：机器学习基础技术 11

2.1 单变量回归 11

2.2 多变量线性回归 18

2.2.1 业务理解 18

2.2.2 数据理解和数据准备 18

2.2.3 模型构建与模型评价 21

2.3 线性模型中的其他问题 30

2.3.1 定性特征 30

2.3.2 交互项 32

2.4 小结 34

第3章 逻辑斯蒂回归与判别分析 35

3.1 分类方法与线性回归 35

3.2 逻辑斯蒂回归 36

3.2.1 业务理解 36

3.2.2 数据理解和数据准备 37

3.2.3 模型构建与模型评价 41

3.3 判别分析概述 46

3.4 多元自适应回归样条方法 50

3.5 模型选择 54

3.6 小结 57

第4章 线性模型中的高级特征选择技术 58

4.1 正则化简介 58

4.1.1 岭回归 59

4.1.2 LASSO 59

4.1.3 弹性网络 60

4.2 商业案例 60

4.2.1 业务理解 60

4.2.2 数据理解和数据准备 60

4.3 模型构建与模型评价 65

4.3.1 最优子集	65
4.3.2 岭回归	68
4.3.3 LASSO	71
4.3.4 弹性网络	73
4.3.5 使用glmnet进行交叉验证	76
4.4 模型选择	78
4.5 正则化与分类问题	78
4.6 小结	81
第5章 更多分类技术：K最近邻与 支持向量机	82
5.1 K最近邻	82
5.2 支持向量机	84
5.3 商业案例	86
5.3.1 业务理解	86
5.3.2 数据理解和数据准备	87
5.3.3 模型构建与模型评价	92
5.3.4 模型选择	98
5.4 SVM中的特征选择	100
5.5 小结	101
第6章 分类回归树	103
6.1 本章技术概述	103
6.1.1 回归树	104
6.1.2 分类树	104
6.1.3 随机森林	105
6.1.4 梯度提升	106
6.2 商业案例	106
6.2.1 模型构建与模型评价	107
6.2.2 模型选择	121
6.2.3 使用随机森林进行特征选择	121
6.3 小结	123
第7章 神经网络与深度学习	124
7.1 神经网络介绍	124
7.2 深度学习简介	128
7.3 业务理解	131
7.4 数据理解和数据准备	132
7.5 模型构建与模型评价	136
7.6 深度学习示例	139
7.6.1 H2O背景介绍	139
7.6.2 将数据上传到H2O平台	140
7.6.3 建立训练数据集和测试 数据集	141
7.6.4 模型构建	142
7.7 小结	146
第8章 聚类分析	147
8.1 层次聚类	148
8.2 K均值聚类	149
8.3 果瓦系数与围绕中心的划分	150
8.3.1 果瓦系数	150
8.3.2 PAM	151
8.4 随机森林	151
8.5 业务理解	152
8.6 数据理解与数据准备	152
8.7 模型构建与模型评价	155
8.7.1 层次聚类	155

8.7.2 K均值聚类	162
8.7.3 果瓦系数和PAM	165
8.7.4 随机森林与PAM	167
8.8 小结	168
第9章 主成分分析	169
9.1 主成分简介	170
9.2 业务理解	173
9.3 模型构建与模型评价	176
9.3.1 主成分抽取	176
9.3.2 正交旋转与解释	177
9.3.3 根据主成分建立因子得分	178
9.3.4 回归分析	178
9.4 小结	184
第10章 购物篮分析、推荐引擎与序列分析	185
10.1 购物篮分析简介	186
10.2 业务理解	187
10.3 数据理解和数据准备	187
10.4 模型构建与模型评价	189
10.5 推荐引擎简介	192
10.5.1 基于用户的协同过滤	193
10.5.2 基于项目的协同过滤	194
10.5.3 奇异值分解和主成分分析	194
10.6 推荐系统的业务理解	198
10.7 推荐系统的数据理解与数据准备	198
10.8 推荐系统的建模与评价	200
10.9 序列数据分析	208
10.10 小结	214
第11章 创建集成多类分类	215
11.1 集成模型	215
11.2 业务理解与数据理解	216
11.3 模型评价与模型选择	217
11.4 多类分类	219
11.5 业务理解与数据理解	220
11.6 模型评价与模型选择	223
11.6.1 随机森林	224
11.6.2 岭回归	225
11.7 MLR集成模型	226
11.8 小结	228
第12章 时间序列与因果关系	229
12.1 单变量时间序列分析	229
12.2 业务理解	235
12.3 模型构建与模型评价	240
12.3.1 单变量时间序列预测	240
12.3.2 检查因果关系	243
12.4 小结	249
第13章 文本挖掘	250
13.1 文本挖掘框架与方法	250
13.2 主题模型	252
13.3 业务理解	254
13.4 模型构建与模型评价	257
13.4.1 词频分析与主题模型	257
13.4.2 其他定量分析	261
13.5 小结	267

第14章 在云上使用R语言 268

14.1 创建AWS账户 269

14.1.1 启动虚拟机 270

14.1.2 启动Rstudio 272

14.2 小结 274

附录 R语言基础 275

• • • • • ([收起](#))

[精通机器学习_下载链接1](#)

标签

机器学习

R

计算科学

CS

等一等

TP自动化技术、计算机技术

GRANDSTRATEGY

评论

<https://github.com/greatabel/RStudy/tree/master/02Mastering%20Machine%20Learning%20with%20R>

每一章的案例讲解很清晰，代码蛮简洁的~

基本不会推导梳理内容，但是会列出参考文献。比较基础，偏入门

LASSO、惩罚项、SVM，很好，我知道了什么是我不知道的。|
好看的封面&第一章里面超厉害的流程总结让我给它满分（希望我的进度不要止步于此）。以及，逼迫自己“理智地”思考也是有点抖M了

有灵魂的一本书，适合入门，讲述清晰。
如本书副标题所说，“理论与技术的完美并重”，不管是对模型的理论讲解还是对业务逻辑的分析都很清晰明确，颇有洞察。

作为一个研一新生，在读完第二章线性回归之后，发现本书并不像其他机器学习的书一样仅仅介绍模型及预测，而是从数理统计的角度来分析并解释模型，全书要求的数学基础较高，并要大致了解r语言，总体来说值得初学者一读。

[精通机器学习_下载链接1](#)

书评

[精通机器学习_下载链接1](#)