

目录

前言	9
入门篇	12
第 1 章：全书指南.....	13
1.1 Python 编程	13
1.2 数据分析.....	17
1.3 机器学习.....	19
1.4 Kaggle 竞赛.....	30
1.5 Git 代码管理.....	31
1.6 章末小结.....	32
第 2 章：基本环境搭建与配置.....	34
2.1 Windows 下基本环境的搭建与配置.....	34
2.2 MacOS 下基本环境的搭建与配置.....	40
2.3 Ubuntu 下基本环境的搭建与配置.....	46
2.4 Jupyter Notebook 使用简介	53
2.5 PyCharm 使用简介.....	57
2.6 章末小结.....	61
基础篇	63
第 3 章：PYTHON3 编程	64
3.1 Python 环境配置	64
3.2 Python 基本语法	70
3.3 Python 数据类型.....	73
3.4 Python 数据运算	77
3.5 Python 流程控制.....	82
3.6 Python 函数设计	86
3.7 Python 面向对象编程	87
3.8 Python 编程库（包）/模块导入	90
3.9 Python 编程综合实践	93
3.10 章末小结.....	94
第 4 章：PANDAS 数据分析	96
4.1 Pandas 环境配置	97
4.2 Pandas 核心数据结构	101
4.3 Pandas 读取/写入文件数据	104
4.4 Pandas 数据分析的常用功能	116
4.5 Pandas 数据合并	130
4.6 Pandas 数据清洗	132
4.7 Pandas 数据分组与聚合	134
4.8 章末小结.....	136
第 5 章：SCIKIT-LEARN 单机机器学习	138
5.1 Scikit-learn 环境配置	139

5.2	Scikit-learn 无监督学习	143
5.3	Scikit-learn 监督学习模型	167
5.4	Scikit-learn 半监督学习模型	209
5.5	单机机器学习模型的常用优化技巧	217
5.6	章末小结	243
进阶篇		247
第 6 章：PYTORCH/TENSORFLOW/PADDLEPADDLE 深度学习		248
6.1	PyTorch/TensorFlow/PaddlePaddle 环境配置	254
6.2	前馈神经网络	263
6.3	卷积神经网络	275
6.4	残差神经网络	289
6.5	循环神经网络	289
6.6	注意力机制	302
6.7	自动编码器	302
6.8	神经网络模型的常用优化技巧	315
6.9	章末小结	339
第 7 章：PYSPARK-ML 分布式机器学习		342
7.1	PySpark 环境配置	345
7.2	PySpark 分布式数据结构	351
7.3	PySpark 分布式特征工程	357
7.4	PySpark-ML 分布式机器学习模型	369
7.5	分布式机器学习模型的常用优化技巧	379
7.6	章末小结	385
实践篇		388
第 8 章：KAGGLE 竞赛实践		389
8.1	Titanic 罹难乘客预测	390
8.2	Ames 房产价值评估	398
8.3	Twitter 短文本分类	409
8.4	CIFAR-100 图像识别	416
8.5	章末小结	426
第 9 章：GIT 代码管理		428
9.1	Git 本地环境搭建	429
9.2	Git 远程仓库配置	432
9.3	Git 基本指令	438
9.4	Git 分支管理	443
9.5	贡献 Git 项目	449
9.6	章末小结	453
后记		456