

Python全栈数据工程师养成攻略(视频讲解版)



[Python全栈数据工程师养成攻略\(视频讲解版\)_下载链接1](#)

著者:张宏伦

出版者:人民邮电出版社

出版时间:2017-11

装帧:平装

isbn:9787115468697

本书首先介绍了数据工程和Python语法，随后讲解如何获取和存储数据，并实现简单的静态可视化。文本作为一种极其重要的数据类型，也单独列出一章进行讨论。之后读者将学习到关于Web建站的一些基础和进阶知识，并基于多种工具完成丰富的动态可视化。最后，这本书选取了机器学习和深度学习两大热门领域的核心内容，为读者进一步实现数据价值的深度分析和挖掘打下坚实基础。

作者介绍:

张宏伦，上海交通大学在读博士。

目录: 第1章 写在前面 1
1.1 数据工程和编程语言 1
1.1.1 如何玩转数据 1

1.1.2 关于编程语言	3
1.2 带好装备——Python 和 Sublime	4
1.2.1 Python	4
1.2.2 Sublime	5
1.2.3 运行 Python 代码的方法	6
1.2.4 Hello World	7
1.3 数据结构和常见类型	7
1.3.1 数据的结构	8
1.3.2 数据的类型	8
第 2 章 学会 Python	10
2.1 Python 基础语法	10
2.1.1 Python 的特点	10
2.1.2 中文编码	10
2.1.3 变量	11
2.1.4 注释	14
2.1.5 保留名	14
2.1.6 行和缩进	15
2.1.7 运算符	15
2.1.8 条件	15
2.1.9 循环	16
2.1.10 时间	18
2.1.11 文件	19
2.1.12 异常	19
2.1.13 函数	20
2.1.14 补充内容	20
2.2 实战：西游记用字统计	21
2.2.1 数据	21
2.2.2 目标	21
2.2.3 步骤	21
2.2.4 总结	23
第 3 章 获取数据	24
3.1 HTTP 请求和 Chrome	24
3.1.1 访问一个链接	24
3.1.2 Chrome 浏览器	25
3.1.3 HTTP	27
3.1.4 URL 类型	28
3.2 使用 Python 获取数据	29
3.2.1 urllib2	29
3.2.2 GET 请求	29
3.2.3 POST 请求	30
3.2.4 处理返回结果	30
3.3 实战：爬取豆瓣电影	31
3.3.1 确定目标	31
3.3.2 通用思路	32
3.3.3 寻找链接	32
3.3.4 代码实现	34
3.3.5 补充内容	38
第 4 章 存储数据	40
4.1 使用 XAMP 搭建 Web 环境	40
4.1.1 Web 环境	40
4.1.2 偏好设置	41
4.1.3 Hello World	43
4.2 MySQL 使用方法	44
4.2.1 基本概念	44

- 4.2.2 命令行 44
- 4.2.3 Web 工具 44
- 4.2.4 本地软件 47
- 4.3 使用 Python 操作数据库 49
 - 4.3.1 MySQLdb 49
 - 4.3.2 建立连接 49
 - 4.3.3 执行操作 50
 - 4.3.4 关闭连接 52
 - 4.3.5 扩展内容 52
- 第 5 章 静态可视化 53
 - 5.1 在 R 中进行可视化 53
 - 5.1.1 下载和安装 53
 - 5.1.2 R 语言基础 54
 - 5.1.3 ggplot2 59
 - 5.1.4 R 语言学习笔记 59
 - 5.2 掌握 ggplot2 数据可视化 59
 - 5.2.1 图形种类 59
 - 5.2.2 基本语法 60
 - 5.2.3 条形图 61
 - 5.2.4 折线图 61
 - 5.2.5 描述数据分布 62
 - 5.2.6 分面 62
 - 5.2.7 R 语言数据可视化 62
 - 5.3 实战：Diamonds 数据集探索 63
 - 5.3.1 查看数据 63
 - 5.3.2 价格和克拉 64
 - 5.3.3 价格分布 64
 - 5.3.4 纯净度分布 65
 - 5.3.5 价格概率分布 65
 - 5.3.6 不同切工下的价格分布 65
 - 5.3.7 坐标变换 66
 - 5.3.8 标题和坐标轴标签 66
- 第 6 章 自然语言理解 67
 - 6.1 走近自然语言理解 67
 - 6.1.1 概念 67
 - 6.1.2 内容 67
 - 6.1.3 应用 68
 - 6.2 使用 jieba 分词处理中文 70
 - 6.2.1 jieba 中文分词 70
 - 6.2.2 中文分词 70
 - 6.2.3 关键词提取 72
 - 6.2.4 词性标注 73
 - 6.3 词嵌入的概念和实现 73
 - 6.3.1 语言的表示 73
 - 6.3.2 训练词向量 75
 - 6.3.3 代码实现 75
- 第 7 章 Web 基础 78
 - 7.1 网页的骨骼：HTML 78
 - 7.1.1 HTML 是什么 78
 - 7.1.2 基本结构 78
 - 7.1.3 常用标签 79
 - 7.1.4 标签的属性 82
 - 7.1.5 注释 83
 - 7.1.6 表单 83

- 7.1.7 颜色 84
- 7.1.8 DOM 85
- 7.1.9 HTML5 86
- 7.1.10 补充内容 86
- 7.2 网页的血肉： CSS 86
 - 7.2.1 CSS 是什么 87
 - 7.2.2 基本结构 87
 - 7.2.3 使用 CSS 87
 - 7.2.4 常用选择器 89
 - 7.2.5 常用样式 91
 - 7.2.6 CSS3 94
 - 7.2.7 CSS 实例 97
 - 7.2.8 补充学习 98
- 7.3 网页的关节： JS 99
 - 7.3.1 JS 是什么 99
 - 7.3.2 使用 JS 99
 - 7.3.3 JS 基础 100
 - 7.3.4 补充学习 103
- 第 8 章 Web 进阶 104
 - 8.1 比 JS 更方便的 JQuery 104
 - 8.1.1 引入 JQuery 104
 - 8.1.2 语法 105
 - 8.1.3 选择器 106
 - 8.1.4 事件 107
 - 8.1.5 直接操作 108
 - 8.1.6 AJAX 请求 112
 - 8.1.7 补充学习 113
 - 8.2 实战：你竟是这样的月饼 113
 - 8.2.1 项目简介 113
 - 8.2.2 首页实现 115
 - 8.2.3 月饼页实现 128
 - 8.2.4 项目总结 133
 - 8.3 基于 ThinkPHP 的简易个人博客 134
 - 8.3.1 ThinkPHP 是什么 134
 - 8.3.2 个人博客 134
 - 8.3.3 下载和初始化 134
 - 8.3.4 MVC 135
 - 8.3.5 数据库配置 136
 - 8.3.6 控制器、函数和渲染模板 137
 - 8.3.7 U 函数和页面跳转 139
 - 8.3.8 表单实现和数据处理 141
 - 8.3.9 读取数据并渲染 142
 - 8.3.10 项目总结 145
 - 8.4 基于 Flask 的简易个人博客 146
 - 8.4.1 Flask 是什么 146
 - 8.4.2 项目准备 147
 - 8.4.3 渲染模板 149
 - 8.4.4 操作数据库 150
 - 8.4.5 完善其他页面 152
 - 8.4.6 项目总结 155
- 第 9 章 动态可视化 157
 - 9.1 使用 ECharts 制作交互图形 157
 - 9.1.1 ECharts 是什么 157
 - 9.1.2 引入 Echarts 158

- 9.1.3 准备一个画板 158
- 9.1.4 绘制 ECharts 图形 158
- 9.1.5 使用其他主题 160
- 9.1.6 配置项手册 160
- 9.1.7 开始探索 164
- 9.2 实战：再谈豆瓣电影数据分析 164
 - 9.2.1 项目成果 164
 - 9.2.2 数据获取 164
 - 9.2.3 数据清洗和存储 167
 - 9.2.4 数据分析 168
 - 9.2.5 数据可视化 168
 - 9.2.6 项目总结 171
- 9.3 数据可视化之魅 D3 172
 - 9.3.1 D3 是什么 172
 - 9.3.2 D3 核心思想 172
 - 9.3.3 一个简单的例子 173
 - 9.3.4 深入理解 D3 177
 - 9.3.5 开始探索 180
- 9.4 实战：星战电影知识图谱 181
 - 9.4.1 项目成果 181
 - 9.4.2 数据获取 182
 - 9.4.3 数据分析 182
 - 9.4.4 数据可视化 183
 - 9.4.5 项目总结 184
- 9.5 艺术家爱用的 Processing 185
 - 9.5.1 Processing 是什么 185
 - 9.5.2 一个简单的例子 186
 - 9.5.3 Processing 基础 186
 - 9.5.4 更多内容 189
- 9.6 实战：上海地铁的一天 189
 - 9.6.1 项目成果 189
 - 9.6.2 项目数据 189
 - 9.6.3 项目思路 190
 - 9.6.4 项目实现 190
 - 9.6.5 项目总结 197
- 第 10 章 机器学习 198
 - 10.1 明白一些基本概念 198
 - 10.1.1 机器学习是什么 198
 - 10.1.2 学习的种类 199
 - 10.1.3 两大痛点 202
 - 10.1.4 学习的流程 203
 - 10.1.5 代码实现 205
 - 10.2 常用经典模型及实现 206
 - 10.2.1 线性回归 206
 - 10.2.2 Logistic 回归 206
 - 10.2.3 贝叶斯 207
 - 10.2.4 K 近邻 207
 - 10.2.5 决策树 207
 - 10.2.6 支持向量机 209
 - 10.2.7 K-Means 209
 - 10.2.8 神经网络 210
 - 10.2.9 代码实现 210
 - 10.3 调参比赛大杀器 XGBoost 213
 - 10.3.1 为什么要调参 214

10.3.2 XGBoost 是什么	214
10.3.3 XGBoost 安装	214
10.3.4 XGBoost 模型参数	215
10.3.5 XGBoost 调参实战	216
10.3.6 总结	227
10.4 实战：微额借款用户人品预测	227
10.4.1 项目背景	227
10.4.2 数据概况	228
10.4.3 缺失值处理	228
10.4.4 特征工程	229
10.4.5 特征选择	230
10.4.6 模型设计	231
10.4.7 项目总结	232
第 11 章 深度学习	233
11.1 初探 Deep Learning	233
11.1.1 深度学习是什么	233
11.1.2 神经元模型	234
11.1.3 全连接层	235
11.1.4 代码实现	236
11.2 用于处理图像的 CNN	237
11.2.1 CNN 是什么	238
11.2.2 CNN 核心内容	239
11.2.3 CNN 使用方法	241
11.2.4 CNN 模型训练	242
11.2.5 代码实现	242
11.3 用于处理序列的 RNN	242
11.3.1 RNN 是什么	242
11.3.2 RNN 模型结构	243
11.3.3 LSTM	244
11.3.4 RNN 使用方法	246
11.3.5 代码实现	246
11.4 实战：多种手写数字识别模型	246
11.4.1 手写数字数据集	247
11.4.2 全连接层	248
11.4.3 CNN 实现	252
11.4.4 RNN 实现	253
11.4.5 实战总结	254
第 12 章 数据的故事	256
12.1 如何讲一个好的故事	256
12.1.1 为什么要做 PPT	256
12.1.2 讲一个好的故事	256
12.1.3 用颜值加分	257
12.1.4 总结	258
12.2 实战：有内容有颜值的分享	258
12.2.1 SODA	258
12.2.2 公益云图	260
12.2.3 上海 BOT	262
12.2.4 总结	263
• • • • •	(收起)

[Python全栈数据工程师养成攻略\(视频讲解版\)_下载链接1](#)

标签

Python

数据分析

全栈

实践者解答

豆瓣

py

1806

评论

其实非常不错，一本薄薄的书籍就包括了Python,R,前端，可视化，机器学习等多方面内容。想通过一本书掌握所有内容是不可能的，本书算起了个引导作用吧，让我们知道某些技术是用来干什么的，基本概念有哪些。

模型离不开数据。在同一个训练集上训练不同的模型，在测试集上可以得到不同的性能；同一个模型，使用不同的训练集进行训练之后，在同一个测试集上的表现也会有所差异。即便是一个学习能力很强的模型，如果没有充足而且高质量的训练数据，模型的参数依然无法得到最优化的调整；即便是一份充足而且高质量的训练数据，如果模型的学习能力不够强，依然无法捕捉到输入特征和输出标签之间的关联。即便以上两项条件都满足，我们依然无法保证训练好的模型，在任何测试集上都能取得同样好的性能，因为测试集的组成和质量也是千差万别、参差不齐的。为了在具体的实际应用中取得尽可能好的结果，我们需要准备更好、更充足的训练数据，探索更好更强大的学习模型，并且在各种各样的测试集上评估模型的性能。

讲了些大方向

一本通俗，较全面的关于python在网页设计，机器学习等计算机数据处理方面的了解读物，内容循序渐进并加以自己的理解，和自己的经历结合真实清晰，还有就感觉这就是国内顶尖985大学博士的实力吗？好强

2017年出版的书讲的还是Python2，这书过时了吧……

[Python全栈数据工程师养成攻略\(视频讲解版\) 下载链接1](#)

书评

[Python全栈数据工程师养成攻略\(视频讲解版\) 下载链接1](#)