

# 《Python机器学习经典实例》阅读计划

——图灵大数据与机器学习群（第1期）

**领读人：猫煞**

## 本书特色

通过实际案例的形式，讲述机器学习的基础知识并运用 Python 语言将机器学习领域常用算法加以实践，让读者能够更好地掌握相关知识点。

## 适合读者

对 Python 语言语法有基本了解者即可阅读，但更推荐熟练掌握 Python 语言且对机器学习有兴趣的读者阅读。

**答疑时间安排：每周三晚 20:00—22:00**

**总阅读时长（预估）：一个月（阅读过程中敲代码实践）**

**每天阅读用时：2小时**

图灵社区本书网址：<http://www.ituring.com.cn/book/1894>

图灵阅读计划网址：<https://github.com/BetterTuring/turingWeChatGroups>

## 阅读规划

### 第一部分（第 1 ~ 3 章）

**阅读时长：10~15小时**

#### 重点内容

##### 1. 数据预处理技术

数据预处理是所有数据科学研究的第一步，针对数据的处理过程又比较繁琐和单调，某些形式的数据甚至很难处理，这部分需要读者熟练掌握 Python 处理各类型数据的方法以及充分了解数据本身。

##### 2. 建立分类器

分类器是对数据的进一步处理，对后续的机器学习内容有重要意义。

## 难点内容

1. 数据预处理
2. 分类器与建模

## 补充

- 一定要理解各个概念的**核心思想**，技术细节太难的部分可以根据自身能力适当跳过。
- 参考书籍： [《Python 数据处理》](#)

## 第二部分（第 4 章）

阅读时长：10~15小时

### 重点内容

- 各种针对不同形式数据的聚类算法

### 难点内容

- 算法的数学推到和理论思想

### 补充

- 对于数学能力较普通的读者，建议**第一遍阅读时只掌握算法思想和实现方式**，之后再针对数学推理部分做深入学习。
- 参考书籍： [《图解机器学习》](#)

## 第三部分（第 5 章）

阅读时长：10小时

### 重点内容

- 推荐引擎的相关技术  
综合数据处理、分类器知识和算法知识实现一个推荐引擎

### 难点内容

1. 前两部分知识点的综合运用
2. 数据、分类器、算法的合理搭配

## 补充

- 跟着书上的实例实现推荐引擎后建议**自己再做一些修改**，套用上不同的分类器不同的算法和数据，实现一个其他类型数据的推荐引擎。
- 参考书籍: [《推荐系统》](#)

## 第四部分（剩余所有章节）

阅读时长：50+小时（包括阅读、编写代码、开发调试和自我拓展）

## 重点内容

1. 各个不同领域数据的采集和预处理  
这部分的各个章节因为关注于不同领域的的数据，由多个实战性的项目组成，每个项目的类型和特点天差地别，数据采集方式也大相径庭，所以是一个比较**综合性的挑战**，比较重要。
2. 针对不同数据如何选择算法和分类器以及模型
3. 数据可视化  
这部分内容虽然与机器学习的核心部分技术关系不大，但是对于人机交互部分是极为重要的。

## 难点内容

1. 对不同领域数据的处理
2. 对不同领域数据采取不同的计算操作
3. 数据可视化的各种方法和形式

## 补充

- 可以开始参考各领域的相关书籍、博客、论文等资料进行深入学习。
- 参考书籍
  - [《Keras图像深度学习实战》](#)
  - [《人脸识别原理及算法》](#)
  - [《机器学习Web应用》](#)
  - [《数据可视化实战：使用D3设计交互式图表》](#)

## 其他建议

- 在阅读过程中遇到问题不要过分死抠，多利用搜索引擎查找相关资料，及时在群内寻求指导，切忌闭门造车。
- 对于实在很难理解的知识点，可以酌情跳过，之后再深入学习和推敲。
- 在完成书上的代码实践后，建议自己灵活运用学到的内容对实例进行一些调整修改，将孤立的理论融会贯通。