

Machine learning practices

Prof. Shuyuan Yang, Zhixi Feng

E-mail: syyang@xidian.edu.cn

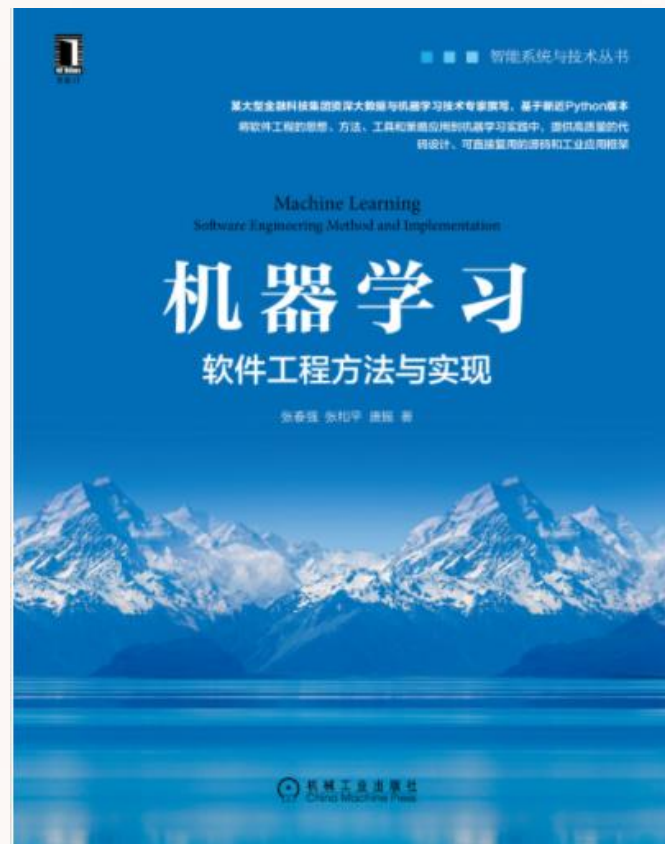
zxfeng@xidian.edu.cn



软件工程

一般来说，软件开发过程包含如下的4项基本活动。

- 1) 软件描述：需求双方定义所要生产的软件需求和约束。
- 2) 软件开发：软件设计与编码。
- 3) 软件有效性验证：检查软件功能与需求是否匹配，排除Bug。
- 4) 软件进化/迭代：软件随变化的需求而进行修改、定制或升级。



机器学习—软件工程方法与实现

- 1) 机器学习建模项目描述：描述目标，明确需求、模型设计和场景、数据、资源等相关约束。
- 2) 建模过程：数据分析、特征处理、算法选用和调参的过程。
- 3) 模型性能评估与报告：评估模型效果是否达到预期和总结报告。
- 4) 模型的迭代和更新：外部环境变化导致数据变化，需要对模型重构或重建。

机器学习中有这样的说法：80%的时间花在数据和特征工程上，20%的时间花在算法模型上。

机器学习项目流程与核心概念

1. 如何定义Y
2. 如何取样X
3. 如何划分数据集
4. 如何选择学习算法
5. 数据分析和处理
6. 特征工程
7. 模型训练与调参
8. 模型评估与报告
9. 模型部署
10. 模型监控
11. 模型重训或重建

机器学习算法8个核心概念

- 损失函数和正则化
- 欠拟合与过拟合、偏差与方差
- 交叉验证
- 数据泄露