第一篇 机器学习基础篇1  
第1 章 线性模型 2  
1．1 概述2  
1．2 算法笔记精华2  
1．2．1 普通线性回归2  
1．2．2 广义线性模型5  
1．2．3 逻辑回归5  
1．2．4 线性判别分析7  
1．3 Python 实战10  
1．3．1 线性回归模型11  
1．3．2 线性回归模型的正则化12  
1．3．3 逻辑回归22  
1．3．4 线性判别分析26

第2 章 决策树 30  
2．1 概述30  
2．2 算法笔记精华30  
2．2．1 决策树原理30  
2．2．2 构建决策树的3 个步骤31  
2．2．3 CART 算法37  
2．2．4 连续值和缺失值的处理42  
2．3 Python 实战43  
2．3．1 回归决策树（DecisionTreeRegressor） 43  
2．3．2 分类决策树（DecisionTreeClassifier） 49  
2．3．3 决策图54

第3 章 贝叶斯分类器 55  
3．1 概述55  
3．2 算法笔记精华55  
3．2．1 贝叶斯定理55  
3．2．2 朴素贝叶斯法56  
3．3 Python 实战59  
3．3．1 高斯贝叶斯分类器（GaussianNB） 61  
3．3．2 多项式贝叶斯分类器（MultinomialNB） 62  
3．3．3 伯努利贝叶斯分类器（BernoulliNB） 65  
3．3．4 递增式学习partial\_fit 方法69

第4 章 k 近邻法 70  
4．1 概述70  
4．2 算法笔记精华70  
4．2．1 kNN 三要素70  
4．2．2 k 近邻算法72  
4．2．3 kd 树73  
4．3 Python 实践74

第5章 数据降维 83  
5．1 概述83  
5．2 算法笔记精华83  
5．2．1 维度灾难与降维83  
5．2．2 主成分分析（PCA） 84  
5．2．3 SVD 降维91  
5．2．4 核化线性（KPCA）降维91  
5．2．5 流形学习降维93  
5．2．6 多维缩放（MDS）降维93  
5．2．7 等度量映射（Isomap）降维96  
5．2．8 局部线性嵌入（LLE） 97  
5．3 Python 实战99  
5．4 小结118

第6章 聚类和EM 算法 119  
6．1 概述119  
6．2 算法笔记精华120  
6．2．1 聚类的有效性指标120  
6．2．2 距离度量122  
6．2．3 原型聚类123  
6．2．4 密度聚类126  
6．2．5 层次聚类127  
6．2．6 EM 算法128  
6．2．7 实际中的聚类要求136  
6．3 Python 实战137  
6．3．1 K 均值聚类（KMeans） 138  
6．3．2 密度聚类（DBSCAN） 143  
6．3．3 层次聚类（AgglomerativeClustering） 146  
6．3．4 混合高斯（GaussianMixture）模型149  
6．4 小结153

第二篇 机器学习高级篇155  
第7章 支持向量机 156  
7．1 概述156  
7．2 算法笔记精华157  
7．2．1 线性可分支持向量机157  
7．2．2 线性支持向量机162  
7．2．3 非线性支持向量机166  
7．2．4 支持向量回归167  
7．2．5 SVM 的优缺点170  
7．3 Python 实战170  
7．3．1 线性分类SVM 171  
7．3．2 非线性分类SVM 175  
7．3．3 线性回归SVR 182  
7．3．4 非线性回归SVR 186

第8章 人工神经网络 192  
8．1 概述192  
8．2 算法笔记精华192  
8．2．1 感知机模型192  
8．2．2 感知机学习算法194  
8．2．3 神经网络197  
8．3 Python 实战205  
8．3．1 感知机学习算法的原始形式205  
8．3．2 感知机学习算法的对偶形式209  
8．3．3 学习率与收敛速度212  
8．3．4 感知机与线性不可分数据集213  
8．3．5 多层神经网络215  
8．3．6 多层神经网络与线性不可分数据集216  
8．3．7 多层神经网络的应用219

第9章 半监督学习 225  
9．1 概述225  
9．2 算法笔记精华226  
9．2．1 生成式半监督学习方法226  
9．2．2 图半监督学习228  
9．3 Python 实战234  
9．4 小结243

第10章 集成学习 244  
10．1 概述244  
10．2 算法笔记精华244  
10．2．1 集成学习的原理及误差244  
10．2．2 Boosting 算法246  
10．2．3 AdaBoost 算法246  
10．2．4 AdaBoost 与加法模型252  
10．2．5 提升树253  
10．2．6 Bagging 算法256  
10．2．7 误差-分歧分解257  
10．2．8 多样性增强259  
10．3 Python 实战260  
10．3．1 AdaBoost 261  
10．3．2 Gradient Tree Boosting 272  
10．3．3 Random Forest 288  
10．4 小结298

第三篇 机器学习工程篇299  
第11章 数据预处理 300  
11．1 概述300  
11．2 算法笔记精华300  
11．2．1 去除唯一属性300  
11．2．2 处理缺失值的三种方法301  
11．2．3 常见的缺失值补全方法302  
11．2．4 特征编码307  
11．2．5 数据标准化、正则化308  
11．2．6 特征选择310  
11．2．7 稀疏表示和字典学习313  
11．3 Python 实践316  
11．3．1 二元化316  
11．3．2 独热码317  
11．3．3 标准化321  
11．3．4 正则化325  
11．3．5 过滤式特征选取326  
11．3．6 包裹式特征选取330  
11．3．7 嵌入式特征选取334  
11．3．8 学习器流水线（Pipeline） 339  
11．3．9 字典学习340

第12 章 模型评估、选择与验证 345  
12．1 概述345  
12．2 算法笔记精华346  
12．2．1 损失函数和风险函数346  
12．2．2 模型评估方法348  
12．2．3 模型评估349  
12．2．4 性能度量350  
12．2．5 偏差方差分解356  
12．3 Python 实践357  
12．3．1 损失函数357  
12．3．2 数据集切分359  
12．3．3 性能度量370  
12．3．4 参数优化387

第四篇 Kaggle 实战篇401  
第13 章 Kaggle 牛刀小试 402  
13．1 Kaggle 简介402  
13．2 清洗数据403  
13．2．1 加载数据403  
13．2．2 合并数据406  
13．2．3 拆分数据407  
13．2．4 去除唯一值408  
13．2．5 数据类型转换410  
13．2．6 Data\_Cleaner 类412  
13．3 数据预处理415  
13．3．1 独热码编码415  
13．3．2 归一化处理419  
13．3．3 Data\_Preprocesser 类421  
13．4 学习曲线和验证曲线424  
13．4．1 程序说明424  
13．4．2 运行结果430  
13．5 参数优化433  
13．6 小结435

全书符号 436