

机器学习实践学习计划

目录

1 学习资料	1
2 内容安排	2
3 学习成果	5

1 学习资料

创造营使用数据酷客平台的机器学习实践课程作为本次计划的学习资源，包含课程视频、视频习题以及案例等。平台课程的内容包含数据预处理、回归模型、分类模型、集成模型、聚类和降维、特征选择和模型选择、机器学习流水线等。本次课程学习计划使用 Python3.x 作为主要的编程语言。

□ 数据酷客网站：

<http://www.cookdata.cn>

□ 机器学习实践课程链接：

http://cookdata.cn/course/course_introduction/39/

2 内容安排

根据学习进度安排表，创造营的小伙伴们可以登录数据酷客平台，学习机器学习实践课程的相关章节内容。我们鼓励并欢迎有兴趣的小伙伴们撰写学习心得笔记。在课程结束后，我们会选择优秀的内容在数据酷客平台发布并和广大的数据酷客用户交流分享。

□ 课程辅导：张嘉田（jiatianz@cookdata.cn）；刘冬（dongl@cookdata.cn）

□ 学习进度安排：

时间	学习内容	学习章节
第一周	机器学习介绍 数据预处理	1.1 机器学习概述 1.2 机器学习常用工具 1.3 Scikit-learn 概览 1.4 实战案例 2.1 数据预处理概述 2.2 缺失值处理 2.3 离群值检测
第二周	数据预处理 回归模型	2.4 标准化 2.5 离散化与特征编码 2.6 实战案例 3.1 回归问题介绍 3.2 线性回归 3.3 线性回归正则化 3.4 稳健性回归

第三周	回归模型	3.5 其它回归方法 3.6 实战案例 4.1 分类问题介绍 4.2 逻辑回归 4.3 K 近邻 4.4 决策树 4.5 朴素贝叶斯
	分类模型	4.6 支持向量机 4.7 实战案例 5.1 聚类问题介绍 5.2 K-Means 聚类 5.3 MeanShift 聚类 5.4 层次聚类 5.5 DBSCAN 5.6 实战案例
第四周	分类模型	4.6 支持向量机 4.7 实战案例 5.1 聚类问题介绍 5.2 K-Means 聚类 5.3 MeanShift 聚类 5.4 层次聚类 5.5 DBSCAN 5.6 实战案例
	聚类模型	5.1 聚类问题介绍 5.2 K-Means 聚类 5.3 MeanShift 聚类 5.4 层次聚类 5.5 DBSCAN 5.6 实战案例
第五周	模型评价与参数调优	6.1 分类指标 6.2 回归指标 6.3 聚类分析指标 6.4 交叉验证 6.5 参数调优
	集成模型	7.1 Bagging
第六周	集成模型	7.2 Boosting 7.3 VotingClassifier 7.4 实战案例
	特征抽取与特征选择	8.1 特征提取与选择概述 8.2 从文本提取特征

第七周	特征抽取与特征选择 降维	8.3 过滤式
		8.4 包裹式
		8.5 嵌入式
		8.6 实战案例
		9.1 降维方法概述
		9.2 线性降维方法
第八周	降维 机器学习流水线	9.3 非线性降维方法
		9.4 实战案例
		10.1 机器学习流水线概述
		10.2 Pipeline
		10.3 FeatureUnion
		10.4 实战案例

3 学习成果

为了激发小伙伴们的学习兴趣、巩固学习的效果，创造营为大家规划了 CapStone 项目。大家可以根据兴趣选择其中一个作为主题，利用所学知识去完成这个 CapStone.

□【必选】撰写课程章节的学习笔记，具体要求如下：

- 易读性. 在观看课程视频的同时，请认真记录老师讲解的内容。在此基础上，组织完整的语句形成学习笔记，杜绝简单的摘抄和罗列. 笔记的内容对于初学者是友好的，容易理解的；
- 完整性. 笔记内容应包含本章节的主要知识点，需要加入自己的阐述和理解，可以做适当的延伸，杜绝机械的重复. 可以使用多种展现形式（如代码、图片）来丰富学习笔记；
- 规范性. 使用外部资料时需要注明引用来源，学习笔记以 Word 格式/Markdown 格式文档提交. 课程助教会评审提交的学习笔记，反馈评审意见. 学习笔记在数据酷客平台发布前，需要根据评审意见进行优化.

□【必选】CapStone 主题：

✓分类主题：在线广告中，点击率（CTR）是评估广告效果的重要指标，随着机器学习技术的不断发展，通过机器学习方法构建自动广告点击预测系统也变得越来越普及. 我们收集了 Avazu 公司一段时间内的广告点击数据，请利用机器学习方法训练 CTR 预估模型，并完成相应的数据分析报告. 报告采用 Jupyter Notebook 的形式，需要包含建模流程文字描述和代码.

报告样例链接：

http://cookdata.cn/note/view_static_note/d5edee1b5ce25bb3afcf4632ca61e517/

✓ 回归主题：汽车是人们常用的交通工具之一，在汽车交易市场中，价格是买家们最关心的问题，借助于机器学习技术可以帮助人们预测汽车的交易价格。我们收集了一家汽车交易平台的交易数据，请利用自己所学知识对数据做分析，并使用机器学习方法建立汽车价格的预测模型，完成相应的数据分析报告。报告采用 Jupyter Notebook 的形式，需要包含建模流程文字描述和代码。

报告样例链接：

http://cookdata.cn/note/view_static_note/\de3dc19bbc87a4cae308bbc7188a7b70/