|  |  |
| --- | --- |
| 机器学习课程大纲 | |
| 机器学习概述 | 2-3课时 |
| KNN | 2课时 |
| 线性回归、Lasso、Ridge【回归算法】 | 2-3课时 |
| 逻辑回归和softmax【分类算法】 | 2-3课时 |
| 特征工程 | 2课时 |
| 决策树 | 2-3课时 |
| 集成学习：随机森林、AdaBoost、GBDT、xgboost、Stacking | 4-6课时 |
| 聚类：K-means | 2-3课时 |
| 朴素贝叶斯 | 2课时 |
| 感知机和SVM | 4课时 |
| 机器学习项目一 | 2-4课时 |
| 机器学习项目二 | 2-4课时 |

说明：课时数目只是预估而且预估的比较紧张，实际上可能会根据实际情况调整，红色部分是比较重要的内容

**第六阶段 机器学习经典算法**

**机器学习概述**

机器学习定义

机器学习、人工智能和深度学习的关系

机器学习基本概念和常用的应用场景

机器学习、数据分析、数据挖掘的区别与联系

机器学习分类

机器学习数据处理流程

**KNN**

KNN算法原理

KNN算法python list实现

KDTree

案例：

KNN伪代码

基于鸢尾花数据的KNN案例

**回归算法**

线性回归算法

多项式扩展

正则化、Lasso&Ridge

机器学习调参

案例：

Linear伪代码

基于解析式的Linear回归python实现

基于波士顿房屋价格的线性回归和多项式扩展

**梯度下降**

梯度下降算法原理

BGD\SGD\MBGD

案例：梯度下降python实现

**Logistic&Softmax**

Odds几率

Logistic原理

Softmax原理

案例：

基于乳腺癌数据的Softmax分类

基于信贷评估数据的Logistic分类

**特征工程**

了解特征工程在机器学习当中的重要性

特征预处理

特征提取

特征选择和特征的降维

案例：

缺省值填充

归一化、标准化

字典数据提取、OneHot、TF-IDF

Jieba分词

PCA降维和LDA降维

**决策树**

信息熵

决策树原理与构建

决策树可视化

案例：

信息熵与不纯度python实现

Python实现决策树构建过程

基于鸢尾花数据的决策树案例及其可视化

**集成学习**

集成学习思想

Bagging、随机森林

Adaboost算法原理

GBDT算法原理

XGBoost算法原理

Stacking算法原理

案例：

Bagging思想python实现

随机森林案例

Boosting思想python实现

Adaboost思想python实现

GBDT思想python实现

基于鸢尾花数据的stacking分类案例

基于时间和电压数据的stacking回归案例

**SVM**

感知器模型

硬间隔SVM

软间隔SVM

核函数

SMO

SVR

案例：

感知器模型python实现

基于鸢尾花数据的不同分类器比较

不同核函数SVM比较

**贝叶斯**

朴素贝叶斯算法原理

贝叶斯网络

案例：

基于鸢尾花数据的朴素贝叶斯案例

**Kmeans**

1. means算法原理

K-means++

K-means||

Mini batch K-means

聚类评估指标

案例：

Python实现K-means伪代码

基于鸢尾花数据的K-means案例

K-means肘方法案例

**天池幸福感预测项目**

项目背景解读及数据解读

数据预处理及特征工程

模型选型及构建

结果预测及提交

**天池金融风控项目**

项目背景解读及数据解读

数据预处理及特征工程

模型选型及构建

结果预测及提交