

机器学习业务实践之路

课程2：机器学习实验工具PAI介绍

阿里云 李博（傲海）



1. 机器学习实验工具概述

2. 企业级机器学习工具

3. 机器学习PAI介绍

机器学习工具概述

单机版机器学习软件：

- R Studio
- Matlab

分布式机器+开源架构：

- Spark MLlib

企业级机器学习云服务：

- AWS ML
- 阿里云PAI

企业级机器学习工具

1. 是否支持大规模数据的计算

- 特征量
- 数据量

2. 是否包含丰富的机器学习算法

- 回归算法：逻辑回归、线性回归。。。
- 分类算法：SVM、随机森林
- 聚类算法：DBSCAN、KMEANS
- 深度学习框架：Tensorflow、Caffe、MXNet

3. 是否提供业务相关服务

- 离线调度
- 模型在线服务

机器学习PAI介绍

算法丰富

提供100余种算法组件，覆盖回归、分类、聚类、文本分析、关系挖掘等算法

深度学习（GPU）

平台支持已经支持Tensorflow、MXNet、Caffe等行业主流深度学习框架，支持底层GPU集群的多卡灵活调用。

可视化操作界面

机器学习平台提供可视化的操作界面，通过拖拉拽的方式拖动算法组件拼接成实验，操作流程类似于搭积木。

一站式服务

提供一站式的服务体验，数据的清洗、特征工程、机器学习算法、评估、在线预测以及离线调度都可以在平台上一站式使用。

PAI用法介绍

数据上传：

- 小于20M通过IDE环境上传
- 大于20M通过 Maxcompute tunnel工具上传

案例的使用：

- 公共数据集
- 内置案例模板

实验的搭建：

- 如何搭建实验
- 如何调试参数

相关服务：

- 社区发帖
- 离线调度
- 在线预测

■ 算法组件推荐

更多

■ PAI产品概述
■ 新手入门
■ 算法组件介绍

更多

■ F&Q
■ 客服在线
■ 常用算法使用案例

更多

基础

基础



新建空白实验

基础



人口普查统计案例

结合人口普查数据搭建实验，统计学历和收入的关系。

2094位用户

基础



心脏病预测案例

从模板创建
查看文档

包括数据预处理、特征工程、模型训练和预测等一套机器学习流程。

1093位用户

基础



【图算法】金融风控实验

利用图算法，针对个人信用，解决金融行业的风控问题。

676位用户

基础



协同过滤做商品推荐

本实验利用协同过滤算法搭建了一套购物推荐流程。

773位用户

基础



农业贷款预测的回归算...

通过回归算法建立模型，预测农业贷款的发放。

393位用户

基础



【文本分析】新闻分类

通过主题模型实现了整个文本分类的流程。

598位用户

基础



【在线预测】中生成...

本实验主要是展示平台在线预测能力，通过中学生的在校园行为预测期末成绩以及对于成绩的关

358位用户

基础



用户流失数据建模demo

分别采用逻辑回归和随机森林两个算法建立用户流失模型并评估两个模型好坏

63位用户

加载更多

相关资料

推荐实验环境：机器学习PAI <https://data.aliyun.com/product/learn>

PAI快速入门： https://help.aliyun.com/document_detail/30350.html

我的个人微信公众号（与我交流）：凡人机器学习

为了无法计算的价值 |  阿里云

