**《从机器学习到大模型实战》培训**

1. **讲师介绍**

邹伟 博士，睿客邦创始人，华东建筑设计研究院研究员、山东交通学院客座教授、硕士生导师，南昌航空大学双师型教师、中国软件行业协会专家委员、上海市计划生育科学研究所特聘专家、天津大学创业导师、中国医药教育协会老年运动与健康分会学术委员、《聊城大学学报》编委；睿客邦当前进入发展第4年，已经与全国十多所高校建立了AI联合实验室，完成和在研50多个AI工业项目，广泛应用于医疗、交通、农业、气象、银行、电信等多个领域。合作企业有：中国电信、中国移动、CSDN，中国建设银行等。在国内外期刊会议发表论文10 余篇，获得国家发明专利1 项，著书1本，译书6本。

1. **培训时间**

**初定 6月2日、3日**

1. **学员要求**

**有Python基础，对AI技术感兴趣，自带笔记本电脑**

1. **培训内容**

|  |  |  |
| --- | --- | --- |
| **模块** | **主要内容** | **代码和案例实践** |
| **第一天：**  **Python与机器学习** | **Python基础学习**   * 解释器Python2.7/3.x、IDE：Anaconda/Pycharm * Python常用库介绍 * **常用的模型和应用** * **AI项目的实战架构流程** | * 公路堵车概率模型Nagel-Schreckenberg实现 * 负二项分布与比赛胜率 * 本福特定律 * 蝴蝶效应：Lorenz系统的三维绘制 |
| **回归分析**   * 线性回归 * Logistic/Softmax回归 * 广义线性回归 * L1/L2正则化 * 梯度下降算法：BGD与SGD * 特征选择与过拟合 | **案例：数据预处理与特征工程**   * 回归算法儿童身高预测 * 股票数据的特征提取和应用 * 泰坦尼克号乘客缺失数据处理和存活率预测 * 广告投入与销售额回归分析 * 鸢尾花数据集的分类 |
| **决策树和随机森林**   * 熵、联合熵、条件熵、KL散度、互信息 * 最大似然估计与最大熵模型 * ID3、C4.5、CART详解 * 决策树的正则化 * 预剪枝和后剪枝 * Bagging Boosting AdaBoost GBDT XGBoost * 随机森林 | **案例：用户行为分析**   * 根据三围信息判断性别 * 葡萄酒数据集的决策树/随机森林分类 * 泰坦尼克乘客存活率估计 * **信用卡用户欺诈检测** * **用户离网分析预测** * **网络运维中吞吐量估计与误差分析** |
| **SVM**   * 线性可分支持向量机 * 软间隔 * 损失函数的理解 * 核函数的原理和选择 * SMO算法 * 支持向量回归SVR * 多分类SVM | **案例：用户分类和聚类**   * 支持向量机手写数字图像识别 * 原始数据和特征提取 * 调用开源库函数完成SVM * 数字图像的手写体识别 * MNIST手写体识别 * SVR用于时间序列曲线预测 * SVM、Logistic回归、随机森林三者的横向比较 * 垃圾邮件分类 |
| **集成学习**   * 个体与集成 * Boosting * 随机森林 * Bagging * GBDT和XGBoost * Stacking | **案例：用户画像**   * 使用集成学习构建初级学习器和次级学习器。 * RF和GBDT实现乳腺癌和鸢尾花的分类 * 内容画像构建（搭建基础、关键词提取、注意力机制） * 用户画像（用户画像的架构、扩展、排序召回） |
| **总结** | **不同模型的使用场景总结**  **各种机器学习模型的调参经验分享** |  |
| **第二天：**  **推荐系统及大模型** | **神经网络结构**   * 滤波器，卷积池化，激活函数，反向传播 * 目标分类与识别、目标检测与追踪AlexNet、VGGNet、GoogleLeNet Inception-V3/V4 * ResNet、DenseNet * **其他视频处理算法简要介绍** | **案例：视频图片分析案例**   * 数字图片分类 * 卷积核与特征提取以图搜图 * 人证合一 * 卷积神经网络调参经验分享 * 睿客狗脸识别案例 * 气象设备海量雨量筒图片识别 * 睿客识云 |
| **经典推荐算法**  隐特征  LFM、SVD、FunkSVD  协同过滤:user-CF/Item-CF | **案例：推荐案例1**  基于LR的线上推荐  基于用户协的文本推荐 |
| **深度学习推荐算法**   * 深度 Matching 方法、（MF 召回法、Embedding 与稀疏 ID 类特征、 Item-CF 召回与 Item2Vec、Airbnb 序列召回，LSH，Simhash） * Graph Embedding 与用户行为构建图（（Deepwalk、Node2Vec、LINE、Alias 采样、Graph Embeddi | **案例：推荐案例2**  Item2Vec 实战  基于图深度网络的新闻推荐项目 |
| **ARIMA、时间序列分析**   * 一维卷积、指数平均和滑动平均、Prophet 模型 * MaxPooling 做光滑Attention CNN+LSTM 模型 * Bi-LSTM 双向循环网络 * Keras 函数式 API 与回调函数 | **案例：时间序列预测模型**   * 时间序列算法：自回归、移动平均和整合模型股票走势可视化 * 突变点分析 * 股票预测模型建立模型训练 * 预测与评估 * 生猪价格预测 |
| **ChatGPT**   * 监督微调（SFT）模型、 * 指示学习和提示学习 * 简单提示、小样本提示、基于用户的提示 * 指令微调 * RLLHF技术详解（从人类的反馈中学习） * 聚合问答数据训练奖励模型（RM） * 强化学习微调、PPO、 * **InstructGPT**遵循用户意图使用强化学习方案 * Instruct Learning vs. Prompt Learning * ChatGPT增加增加了Chat属性 * AI 系统的新范式   GPT1-GPT2-GPT3-InstructGPT的-chatGPT的技术关 | **案例：我们如何基于开源大模型完成我们的项目**   * 实战：使用chatGPT打造你的私人聊天助理（revchatGPT）      * 实战：演示提示词技巧，翻译器润色器、JavaScript 控制台 、Excel Sheet      * 实战：网站定制chatgpt-web |

注：

1. 内容可根据学员需求微调，以现场通知为准。
2. 报名赠送全部PPT和案例源代码、实验录播课及实验手册等资源