泰迪科技就业班培训课程大纲

方向一：Hadoop+Spark 大数据开发

|  |  |  |
| --- | --- | --- |
| Linux基础 | 数据库SQL基础 | Java基础 |
| 1.Linux操作系统简介、安装及使用；  2.Linux档案权限与目录配置、档案与目录管理、用户账号管理；  3.文件操作、文本编辑、vim 程序编辑器；  4.Bash简介及功能介绍、Shell Script编程； | 1.数据库概述、数据库系统、管理系统；  2.MySQL安装配置、基本操作、MySQL连接工具的使用；  3.数据库基础知识概述：数据库、表结构、字段及其数据类型、运算符；  4.SQL概述,数据库及表的创建基本操作；  5.增删改查语句,使用Java程序进行增删改查。 | 1.Java基础程序设计：数据类型、运算符、表达式、判断循环语言；  2.数组定义与使用、方法定义与重载；  3.面向对象（基础）；  4.面向对象（高级）；  5.异常概念与编程；包概念与导入；泛型概念与编程应用；  6.Java类集：Collection接口主要操作方法：List、Set、Map常用子类与使用，集合操作与遍历方法；  7.多线程编程及应用；  8.Java EE编程实战；JavaEE基础，HTML，JavaScript，Java Web，Tomcat，JSP |

|  |  |  |
| --- | --- | --- |
| Hadoop编程开发 | Hive大数据分析 | HBase数据快速读写 |
| 1.Hadoop简介、架构、原理；  2.集群配置及安装（JDK、SSH）；  2. Hadoop IDE开发环境配置（Eclipse配置）；  3. Hadoop MapReduce实战：单词计数；  4. Hadoop命令（hadoop fs , hadoop job, Hadoop jar）；  5.分布式文件系统HDFS Java API实战：创建目录，上传，下载，删除；  5. Hadoop基础编程（Mapper、Reducer、Driver）；  6. Hadoop高级编程（Combiner／Partitioner原理实战及自定义，FileInputFormat、FileOutputFormat）；  8. Hadoop高级编程（自定义键值对）；  9. Hadoop高级编程（开发环境直接运行MapReduce程序）；  10. Hadoop案例1：基于KNN的鸢尾花分类预测；  11. Hadoop案例2：基于Kmeans的客户价值分析。 | 1. Hive产生背景、概念，Hive架构、原理；  2. Hive安装、配置；  3.Hive与RDBMS区别，Hive应用场景；  4. Hive HQL命令，数据库，表创建；数据导入、导出，HQL查询；  5.Hive内置函数，自定义函数（UDF）；  6.Hive与Java交互；  7.基于Hive的航空公司客户价值分析项目案例。 | 1.HBase简介、架构原理；  2.HBase集群安装配置（Zookeeper简介、Zookeeper安装配置、配置文件解读）；  3.HBase架构与数据模型（组件原理功能、Rowkey设计、Schema设计）；  4.HBas命令行（hbase shell）表操作；  5.HBase开发环境配置；  6.Java API操作HBase删/建表、HBase增删改查；  7.HBase与Hadoop Mapreduce交互（HBase to HDFS， HDFS to HBase，HBase to HBase）；  9.基于HBase的冠字号查询系统（案例背景、架构、数据处理、建表、导入表数据实现）；  10.基于HBase的冠字号查询系统Java Web实现。 |

|  |  |
| --- | --- |
| Spark大数据分析与编程 | 大数据实战案例 |
| 1.Spark 简介（Spark简介、应用场景）;  2.Spark安装配置（资源管理器YARN／Spark Standalone）；  3. Spark整合Hive及传统数据库；  3.原理与架构（组件功能、RDD原理）；  4.Spark编程基础（Scala及编程简介）；  5.Spark Transformation/Action编程；  6.Spark Intellij IDEA开发环境配置；  7.基于Spark ALS算法的电影推荐系统（案例背景、系统架构）；  8.基于Spark ALS算法的电影推荐系统（Spark Shell简单实现& Java Web远程调用Spark集群实现）； | 一、基于Hadoop、Spark的法律网智能推荐  1.法律网案例背景，架构设计；  2.数据预处理：异常数据探索、数据剔除；  3.系统数据预处理实现；  4.推荐算法简介、评估与选型；  5.系统模型建立、预研、调优；  6.系统模型完整实现。  二、广电大数据用户营销  1、案例背景、架构设计、数据探索  2、数据探索分析及数据预处理  3、SVM预测用户是否挽留  4、用户画像实现  三、广告流量作弊识别  1、案例背景、架构设计、探索分析  2、预处理广告流量检测数据  3、使用决策树算法进行作弊流量识别  4、使用随机森林算法进行作弊流量识别  四、P2P信用贷款风险预测  1、分析P2P信贷行业的风险控制、了解某P2P平台数据情况、用户逾期预测的步骤与流程  2、探索P2P信贷用户逾期的相关因素  3、P2P信贷用户数据预处理  4、构建用户逾期还款预测模型 |

方向二：Python+TensorFlow 大数据挖掘

|  |  |
| --- | --- |
| Python编程基础 | Python数据分析与应用 |
| 1 准备工作  1.1 认识Python  1.2 搭建Python环境  1.3 安装PyCharm并创建一个应声虫程序  2 Python基础知识  2.1 掌握Python固定语法  2.2 创建字符串变量并提取里面的数值  2.3 计算圆形的各参数  3 Python数据结构  3.1 创建一个列表（list）并进行增删改查操作  3.2 转换一个元组（tuple）并进行取值操作  3.3 创建一个字典（dict）并进行增删改查操作  3.4 将两个列表转换为集合（set）并进行集合运算  4程序流程控制语句  4.1实现考试成绩划分  4.2实现一组数的连加与连乘  4.3使用冒泡排序法排序  4.4实训（猜数字游戏）  5 函数  5.1自定义函数实现输出方差  5.2使用匿名函数添加列表元素  5.3存储并导入函数模块  6 面向对象  6.1 认识面向对象编程  6.2 创建Car类  6.3 创建Car对象  6.4 迭代Car对象  6.5 产生Land\_Rover对象（子类）  7 文件基础  7.1 认识文件  7.2 读取txt文件中的数据  7.3 保存数据为csv格式文件  7.4 认识os模块 | 1 Python数据分析概述  1.1认识数据分析  1.2熟悉Python数据分析的工具  1.3安装Python3的Anaconda发行版  1.4掌握Jupyter Notebook常用功能  2 NumPy数值计算基础  2.1认识NumPy数组对象ndarray  2.2认识NumPy矩阵与通用函数  2.3利用NumPy进行统计分析  3 Matplotlib数据可视化基础  3.1了解绘图基础语法与常用参数  3.2分析特征间的关系  3.3分析特征内部数据分布与分散状况  4 Pandas统计分析基础  4.1读写不同数据源的数据  4.2掌握DataFrame的常用操作  4.3转换与处理时间序列数据  4.4使用分组聚合进行组内计算  4.5创建透视表与交叉表  5 使用Pandas进行数据预处理  5.1合并数据  5.2清洗数据  5.3标准化数据  5.4转换数据  6使用scikit-learn构建模型  6.1使用sklearn转换器处理数据  6.2构建并评价聚类模型  6.3构建并评价分类模型  6.4构建并评价回归模型 |

|  |  |
| --- | --- |
| Linux基础 | 数据库SQL基础 |
| 1.Linux操作系统简介、安装及使用；  2.Linux档案权限与目录配置、档案与目录管理、用户账号管理；  3.文件操作、文本编辑、vim 程序编辑器；  4.Bash简介及功能介绍、Shell Script编程； | 1.数据库概述、数据库系统、管理系统；  2.MySQL安装配置、基本操作、MySQL连接工具的使用；  3.数据库基础知识概述：数据库、表结构、字段及其数据类型、运算符；  4.SQL概述,数据库及表的创建基本操作；  5.增删改查语句,使用Java程序进行增删改查。 |

|  |  |
| --- | --- |
| Python数据可视化 | Python网络爬虫实战 |
| 1 准备工作环境  1.1准备介绍  1.2安装matplotlib、Numpy和Scipy库  1.3安装图像处理工具：Python图像库（PIL）  1.4配置matplotlib参数  2绘制并定制化图表  2.1简介  2.2定义图表类型—柱状图、线形图和堆积柱状图  2.3设置坐标轴长度和范围、线型、属性和格式化字符串  2.4添加图例和注解  3学习更多图表和定制化  3.1简介  3.2设置坐标轴标签的透明度和大小  3.3为图表线条添加阴影  3.4向图表添加数据表  3.5使用subplots(子区)  3.6定制化网格  4创建3D可视化图表  4.1简介  4.2创建3D柱状图  4.3创建3D直方图.  4.4在matplotlib中创建动画  5用图像和地图绘制图表  5.1用PIL做图像处理  5.2在具有其他图形的图表中显示图像  5.3使用Basemap在地图上绘制数据 | 1 Python爬虫环境与爬虫简介  1.1认识爬虫  1.2认识反爬虫  1.3配置Python爬虫环境  2 网页前端基础  2.1认识Python网络编程  2.2认识HTTP协议  3 简单静态网页爬取  3.1实现HTTP请求  3.2解析网页  3.3数据存储  4 常规动态网页爬取  4.1逆向分析爬取动态网页  4.2使用Selenium库爬取动态网页  4.3存储数据至MongoDB数据库  5 模拟登录  5.1使用表单登录方法实现模拟登录  5.2使用Cookie登录方法实现模拟登录  6 终端协议分析  6.1分析PC客户端抓包  7 Scrapy爬虫  7.1认识Scarpy  7.2通过Scrapy爬取文本信息  7.3定制中间件 |

|  |  |
| --- | --- |
| Python机器学习实战 | TensorFlow实战 |
| 1机器学习绪论  1.1 引言  1.2 基本术语  1.3 假设空间  1.4 归纳偏好  2 模型评估与选择  2.1 经验误差与过拟合  2.2 评估方法  2.3 性能度量  2.4 偏差与方差  3 回归分析(Regression Analysis)  3.1 基本形式  3.2 线性模型  3.3 逻辑回归  4 决策树(Decision Tree)  4.1 基本流程  4.2 划分选择  4.3 剪枝  5 人工神经网络(Artificial Neural Network)  5.1 神经元模型  5.2 感知机与多层网络  5.3 误差逆传播  5.4 BP神经网络  6 最近邻算法（KNN）  7 朴素贝叶斯分类(Naive Bayesian)  8 聚类分析(Cluster Analysis)  8.1 聚类任务  8.2 性能度量  8.3 距离计算  8.4 常用聚类算法  9支持向量机(Support Vector Machine)  9.1 间隔与支持向量  9.2 对偶问题  9.3 核函数  9.4 软间隔与正则化 | 1 TensorFlow与深度学习实战  1 TensorFlow安装与入门  1.1 TensorFlow环境搭建  1.2 TensorFlow计算模型性:计算图  1.3 TensorFlow数据模型:张量Tensor  1.4 TensorFlow运行模型:会话  2 TensorFlow数据类型  2.1 常量、变量及其构建  2.2 TensorFlow实现线性回归模型  2.3 操作：TensorFlow实现鸢尾花分类  3 TensorFlow实现多层神经网络  3.1 BP神经网络模型(Back Propagation)  3.2 操作：利用TensorFlow构建BP网络模型实现鸢尾花分类  4 mnist手写数字识别  4.1 图片预处理  4.2 占位符:placeholder  4.3 操作：利用TensorFlow实现SoftMax网络对手写数字分类  5 TensorFlow动态学习速率  5.1 TensorFlow动态学习速率设置  5.2 TensorFlow动态学习速率使用  6 TensorFlow模型保存与调用  6.1 TensorFlow模型保存操作  6.2 TensorFlow模型加载与调用  6.3 操作：softmax网络保存与调用实例 |

|  |  |
| --- | --- |
| 深度学习原理及编程实现 | 10大项目案例 |
| 1卷积神经网络CNN  1.1卷积神经网络(CNN)简介  1.2 CNN关键结构:卷积层与池化层  1.3 经典卷积网络模型: LeNet-5  1.4 图像数据处理  1.5 操作：CNN的TensorFlow实现  2循环神经网络RNN  2.1循环神经网络(RNN)简介  2.2 RNN网络关键结构: 隐层互联  2.3 经典RNN模型: LSTM  2.4 时序数据处理  2.5 操作：RNN的TensorFlow实现  3自然语言处理基础  3.1 自然语言处理简介  3.2 中文分词  3.3 NLP概率图  3.4 文本分类  4词表征（Word Representation）  4.1词表征的背景与应用  4.2 one-hot编码  4.3 word2vec原理及实现  5深度学习挖掘语义特征  5.1词向量组合相似度  5.2 基于循环神经网络（RNN）编码的相似度计算 | **基础案例**  1 数据分析案例：基于水色图像的水质识别  2 数据分析案例：电子商务网站用户行为分析及服务推荐  3 数据采集和分析案例：Python爬虫实践:《流浪地球》豆瓣影评分析  4 数据分析案例：市财政收入分析及预测  5 数据分析案例：广电大数据营销推荐项目实战  6 数据分析和可视化案例：航空公司客户价值分析  7 文本挖掘和可视化案例：基于文本内容的垃圾短信分类  **人工智能案例**  基于卷积神经网络（CNN）的人脸识别系统  **综合案例**  1 京东商城美的空调评论数据情感分析  2 大数据岗位人才招聘信息的分析与挖掘 |