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Python1.2 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ndarray2.2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Scarpy7.2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Tensor1.4 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:placeholder4.3 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CNN1.1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-51.4 图像数据处理1.5 操作：CNN的TensorFlow实现2循环神经网络RNN2.1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: LSTM2.4 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 |