2024-12-20

应用系统体系架构 — 作业12

# 作业要求

请你在大二开发的E-Book系统的基础上，参照上课演示的案例开发程序满足以下要求：

## 使用Spark，在你的E-BookStore中增加如下的功能：

### 将你的系统中所有图书的简介按照图书类型分别存储到多个文本文件中，例如，所有计算机类图书的简介存储在CS.txt中，科幻小说的简介存储在Fiction.txt中。请构建多个这样的文件，作为并行作业的对象。

### 编写一个关键词列表，包含若干单词，例如，["Java","JavaScript","C++","Programming","Star","Robot"]等。

### 编写一个Spark程序，将你的文件读入到内存中变成RDD，通过MR作业来统计所有图书简介中每个关键词出现的次数。

## 完成“作业12 循环神经网络.ipynb”中的要求：

### 你可以使用Keras来构建和训练自己的循环神经网络RNN和LSTM，并通过调整超参数使你的模型的能够准确预测比特币的价格。

### 你应该通过参数优化来提升你的模型的准确性，使其预测误差尽量小。

### 你需要将Notebook运行完的结果保存之后提交，否则你的Notebook中将不包含你的代码和运行结果。

### Notebook对应的.ipynb文件可以用PyCharm进行开发，但是可能会有部分内容无法正确显示，推荐使用Anaconda中的JupyterLab或者JupterNotebook来开发。Anaconda的安装方式可以参考“Jupyter Notebook安装与环境搭建.pdf”

# 提交要求

## 请将你自己编写的源码、脚本和文档压缩后上传，勿压缩整个工程提交，尤其是不要压缩第三方的Jar包。

## 你需要将Notebook运行完的结果保存之后提交，否则你的Notebook中将不包含你的代码和运行结果。

# 评分标准：

## 能够正确配置并运行统计功能，得到正确的结果，在验收时需当面演示。 (3分)

## 能够构建满足要求的RNN和LSTM模型，使得比特币价格预测的误差尽量小 (2分)