实验二：线性回归与LSTM

1. 实验背景

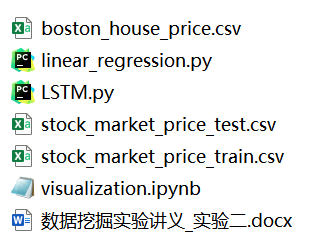
线性回归是利用最小二乘函数对一个或多个自变量和因变量之间关系进行建模的一种回归分析。这种函数是一个或多个称为回归系数的模型参数的线性组合。只有一个自变量的情况称为简单回归，大于一个自变量情况的叫做多元回归。利用线性回归，我们可以预测一组特定数据（如GDP、石油价格和股票价格）是否在一段时期内增长或下降。在本次实验中，我们将使用线性回归预测波士顿房价，进一步理解该算法的原理，并通过可视化工具展示和分析数据集。另外，我们还将搭建一个简单的LSTM网络，实现对股市收盘价格的预测。

1. 实验目的
2. 学习数据集的可视化和分析；
3. 进一步掌握线性回归的原理；
4. 熟悉LSTM的相关知识。
5. 准备工作
6. 安装anaconda，并创建lab2实验环境
7. 在创建环境中，安装至少下列python包：pytorch, jupyter, numpy, sklearn, matplotlib, pandas, seaborn。

（ps：建议提前配置好，机房网络不一定顺畅）

1. 动手实现

我们提供的文件夹目录如下。



boston\_house\_price.csv文件是波士顿房价数据集。visualization.ipynb文件实现数据集的可视化和分析。linear\_regression.ipynb文件使用线性回归预测房价。

Stock\_market\_price\_train和Stock\_market\_price\_test是某股票的价格数据。LSTM.py文件实现股市价格的预测。

你需要做的是补全visualization.ipynb、linear\_regression.py、LSTM.py三个文件的代码。

（1）visualization.ipynb补全

请你补充相关的代码，实现数据集的读取和可视化。根据可视化的结果，探究影响波士顿房价的因素。





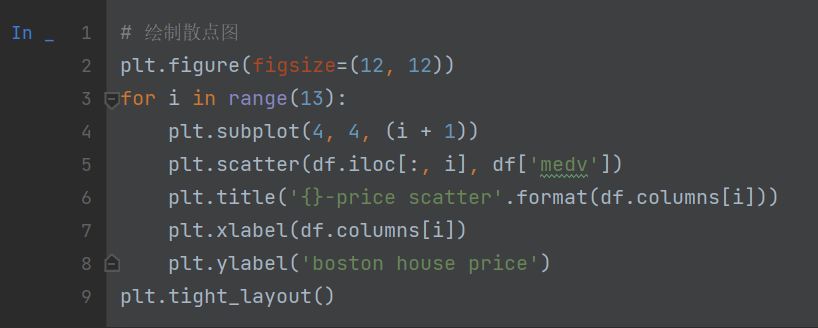
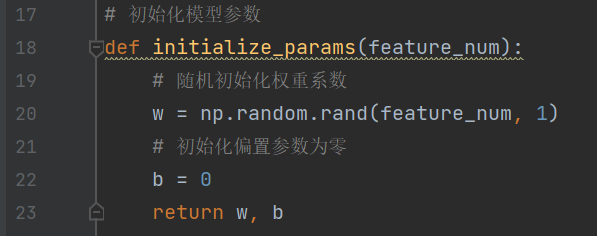


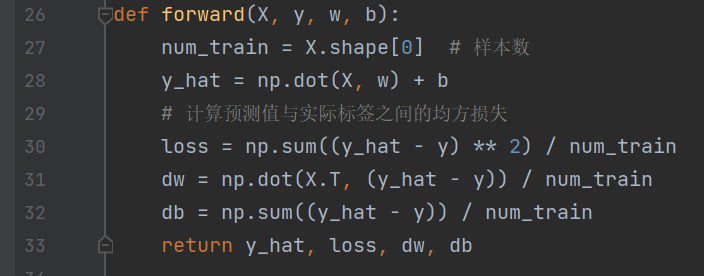
图1 代码

到目前为止，你已经完成了数据集可视化代码的编写。现在你可以编写linear\_regression.py文件完成线性回归算法。

（2）linear\_regression.py补全

现在我们开始完成基于梯度下降的线性回归算法的代码补全。





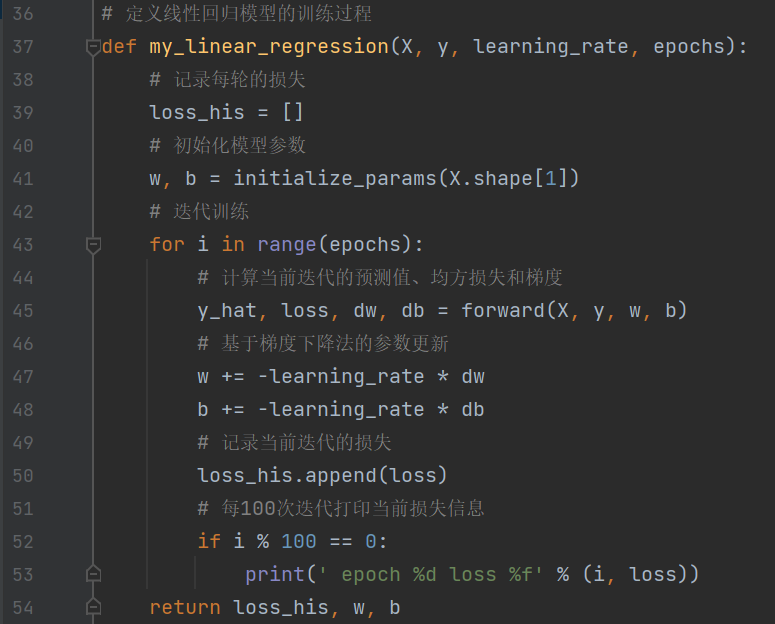
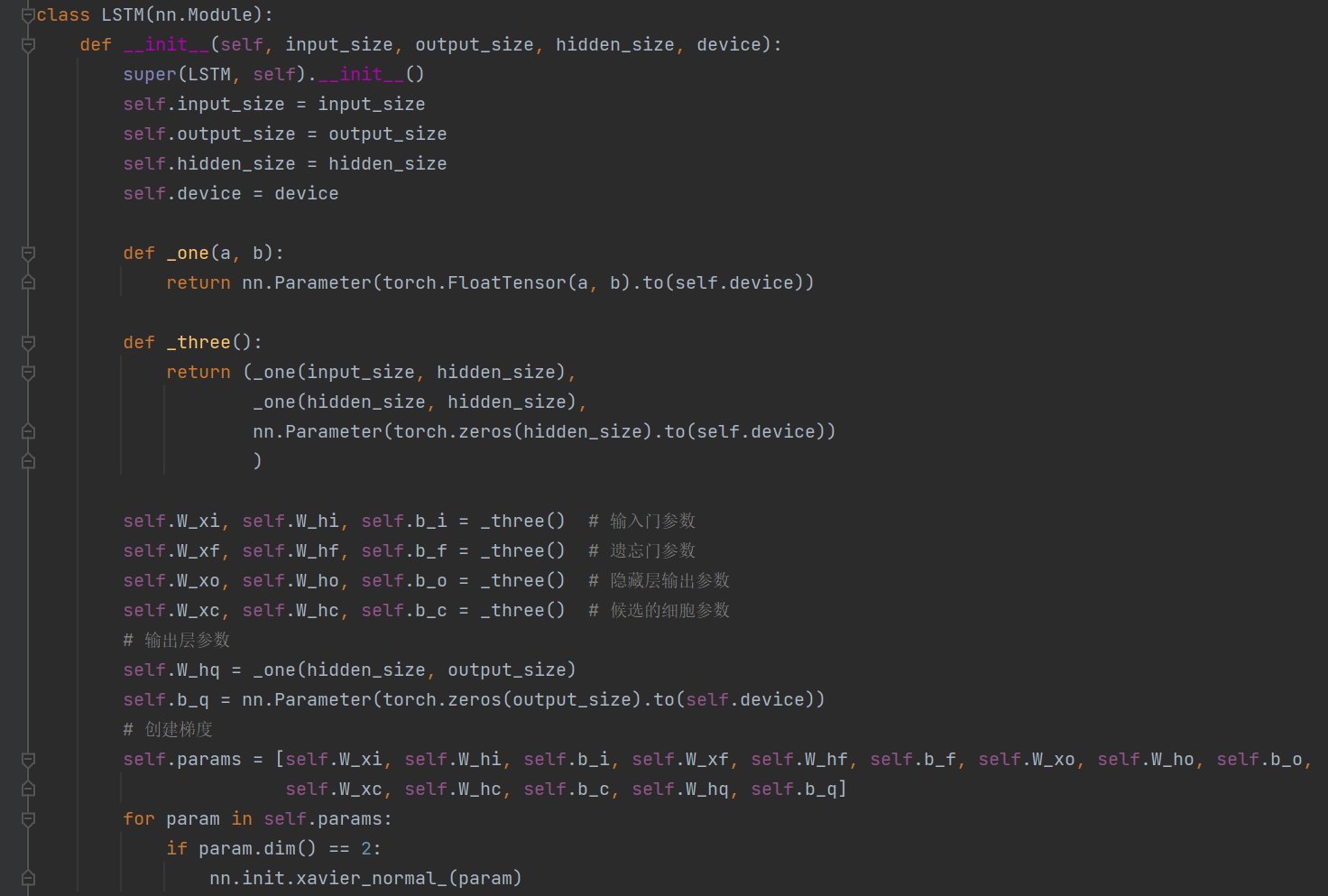


图2 代码

到目前为止，你已经完成了线性回归代码的编写，现在你可以执行代码，查看训练的Loss曲线、在测试集上的Loss和R方系数，观察预测效果。

（3）LSTM.py文件的补全



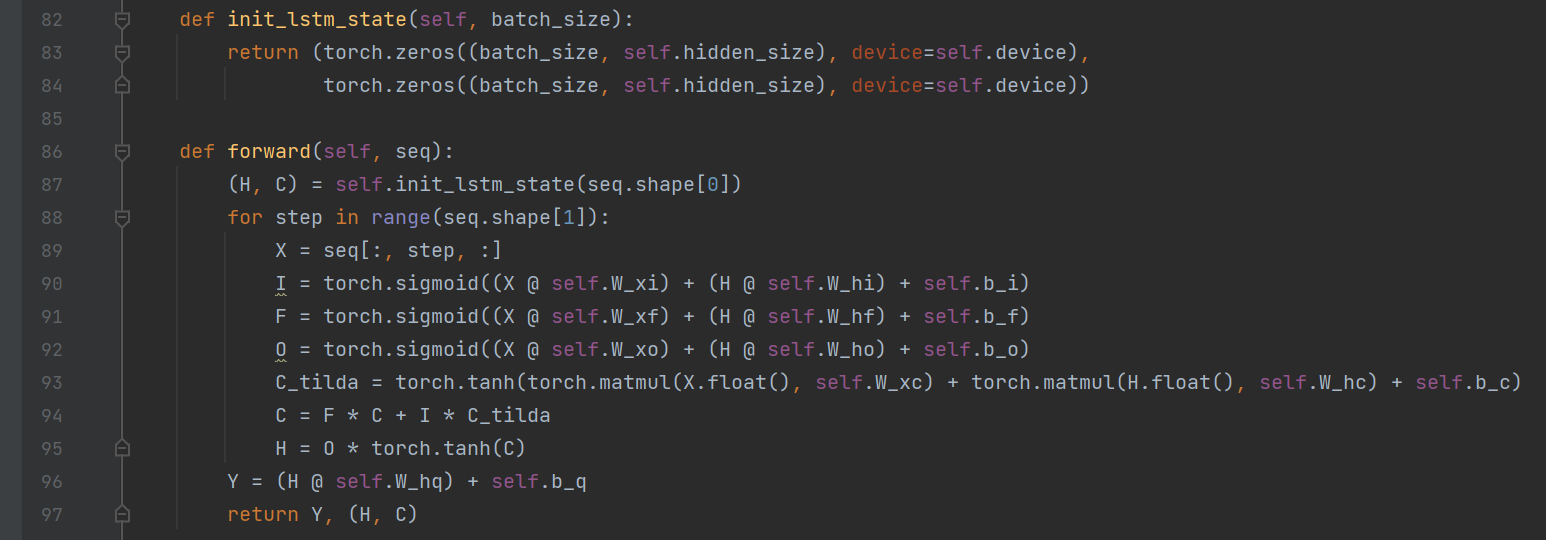


图3 代码

到目前为止，你已经完成了LSTM代码的编写，现在你可以执行代码，查看训练的Loss曲线，和测试集的预测结果。

请整理实验结果并完成实验报告：内容应包括环境配置、核心代码分析、实验过程截图、结果展示、实验体会与收获以及实验反馈与意见（以便后续实验设计调整）。

课外拓展（不要求完成）：进一步探索模型

如果你在下课之前就完成了本次实验，行有余力的同学可以按照下方要求进行拓展，更深刻感受预测算法的区别与优缺点。

1. 你可以使用两层的LSTM预测股票收盘价的变化吗？
2. 你可以利用前几天的收盘价、开盘价、交易额等数据预测下一天的收盘价吗？
3. 评分标准

|  |  |  |
| --- | --- | --- |
| 评分项 | 占比 | 说明 |
| 考勤 | 40% | 特殊情况请提前请假 |
| 代码有效 | 20% | 指代码能成功运行 |
| 实验报告 | 40% | 是否认真完成 |