**四川大学MBA平时作业**

**（** 2020 **至** 2021 **学年第** 二 **学期）**

**课程名称数据模型与决策 授课老师 王緌 成 绩**

**学生姓名 迟丰鑫 学 号 2020225029499 班 级 晚1班**

|  |
| --- |
| **注意事项：**   1. **答案必须统一写在A4纸上，附在此封面下一并交上；** 2. **请完整填写此封面上要求填写的内容；** 3. **请严格按照各科老师要求完成作业；** 4. **字号为宋体小四号，行距为1.5倍行距；** 5. **请独立完成作业，严禁抄袭**   **教师评阅意见：**  **评阅教师签名：**  **年 月 日** |

**基于Pytorch的上市公司财务预警分析**

摘要：根据股市公开披露的财务报表和营运数据，对上市公司进行财务预警，是进行价值投资的有效手段。但这些披露的数据与实际的财务状况之间往往缺乏相关性，让人很难从中发现价值。本文将从数据挖掘的角度，使用开源机器学习库Pytorch，基于长短期记忆网络原理（LSTM），构建分析财务数据与财务预警之间关系的简化模型。

关键字：Pytorch，LSTM，深度学习，财务预警

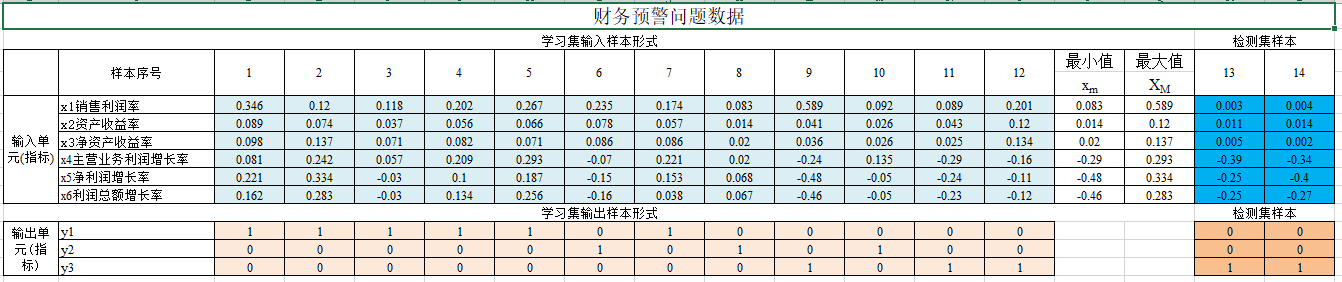
1. 引言

股票市场投资在高回报的同时往往伴随着高风险。要充分挖掘上市公司真实财务状况，在纷繁复杂的公司财报中很难找到线索，并且极易判断失误。但随着大数据和机器学习的普及，使得对相关报表进行分析并找出真实财务状况变得可能。

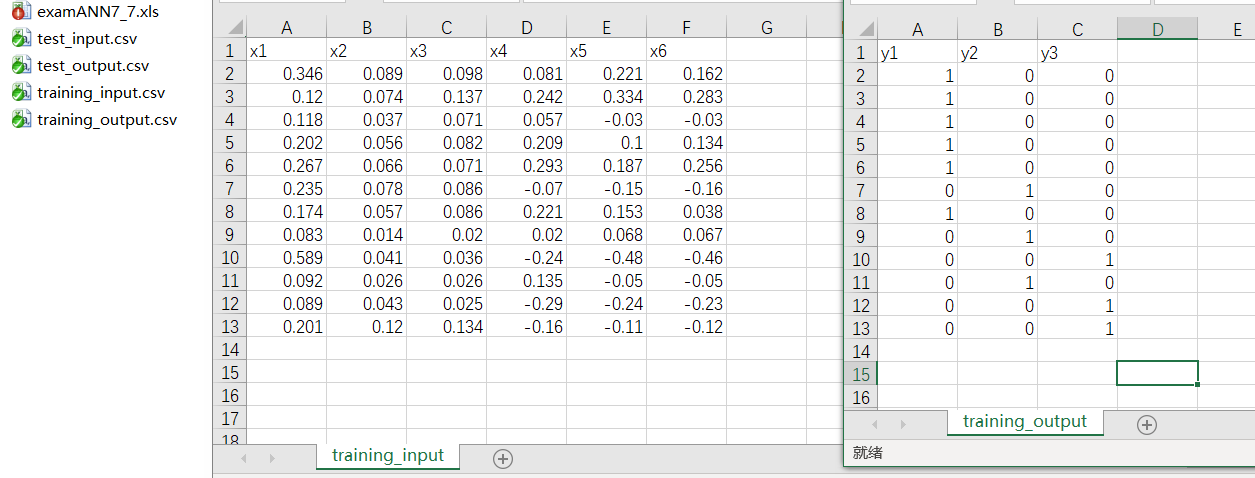
本文将从数据挖掘的角度，通过流行的深度学习库Pytorch，构建一个简易的财务预警分析模型。

1. 需求建模

分析上市公司财务状况，提出与财务预警相关的6个指标，分别为销售利润率、资产收益率、净资产收益率、主营业务利润增长率、净利润增长率、利润总额增长率。依据以上指标，收集到14组样本数据。作为简易模型，将选取12组为训练数据，对模型进行训练，并取两组样本作为测试数据，对最终的模型进行测试。



将以上原始数据整理，分成四个数据文件，分别为训练输入、输出，测试输入、输出文件，保存为csv格式供后续程序导入。



1. Pytorch深度学习框架

PyTorch是一个开源的Python机器学习库，基于Torch，用于自然语言处理等应用程序。

2017年1月，由Facebook人工智能研究院（FAIR）基于Torch推出了PyTorch。它是一个基于Python的可续计算包，提供两个高级功能：1、具有强大的GPU加速的张量计算（如NumPy）。2、包含自动求导系统的深度神经网络。

开发环境

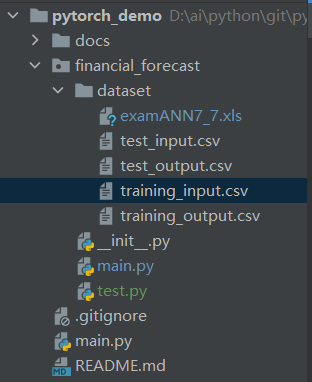
|  |  |  |
| --- | --- | --- |
| 软件和工具 | 描述 | 版本 |
| Python | 运行环境 | 3.8.8 |
| Anaconda | Python多版本管理工具 | Latest |
| Pytorch | 机器学习库 | 1.8.1 |
| PyCharm | 代码编辑器 | 2021.1社区版 |
| Github | 代码管理器 |  |

1. 代码实现

具体编码实现流程如下

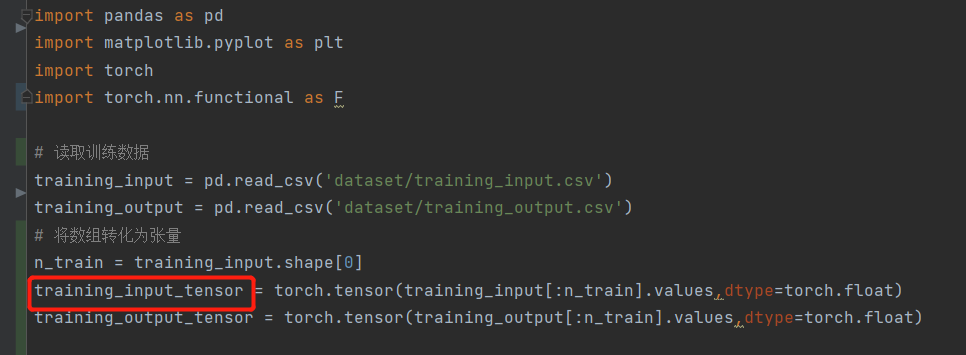
* 1. 读入数据

将四个数据文件放入程序目录下的dataset文件夹

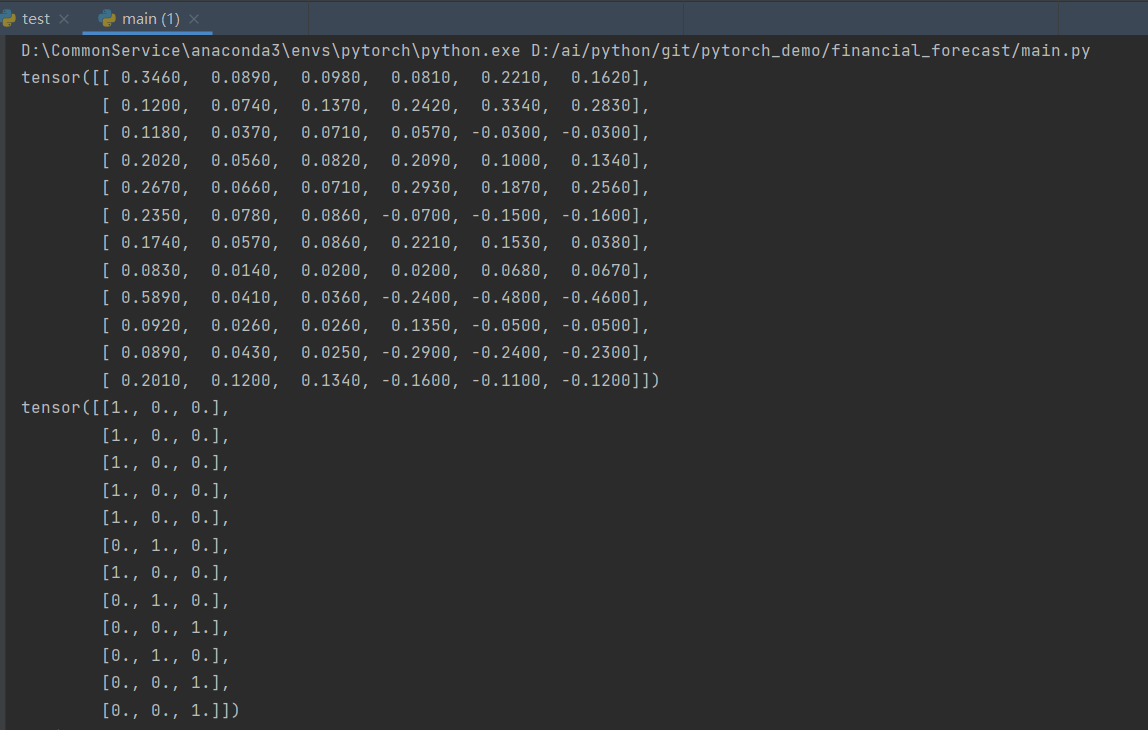


首先将两个训练数据通过pandas组件读入python内存。读入后自动转换为数组形式，并可以通过首行作为列的索引访问，如training\_input.x1

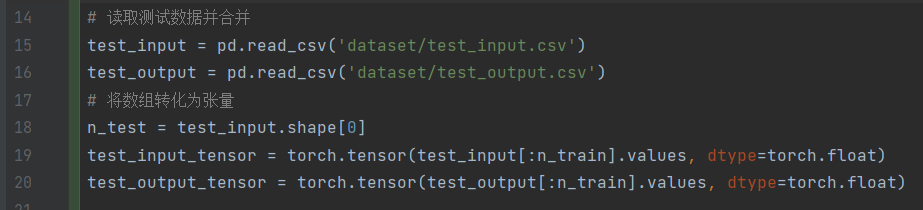
第二步，将数组结构转化为torch中的张量结构，即为两个变量：training\_input\_tensor和training\_output\_tensor



导入后的内存结构如下



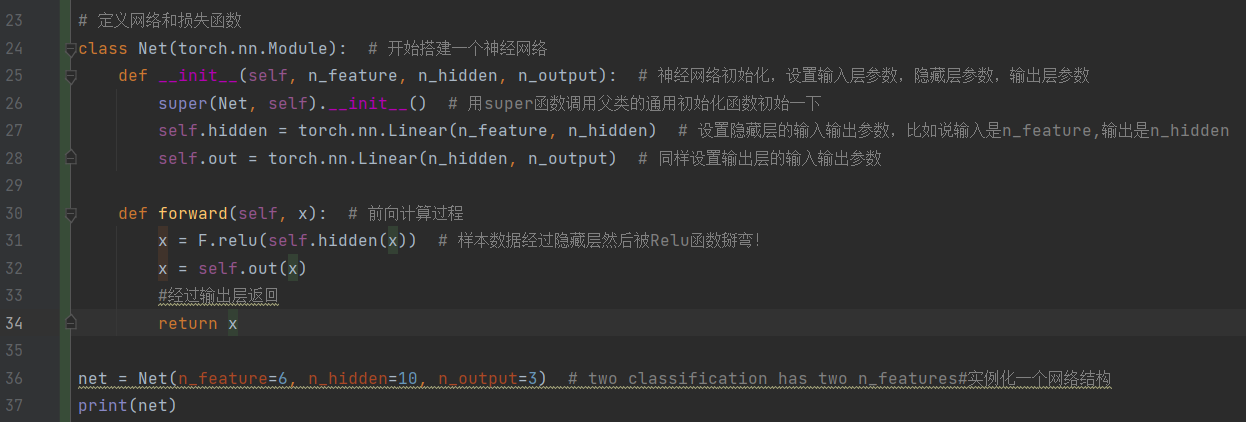
以相同的方式导入测试数据集。



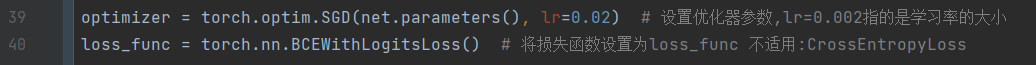
* 1. 定义网络和损失函数

使用torch.nn包中的组件，构建线性回归模型网络Net，并定义前向计算过程。

定义完成后创建网络模型的对象net，在本场景下，根据数据模型6个输入维度和3个输出维度，分别设定n\_feature=6, n\_output=3，同时设定中间隐藏层为10，也就是构建中间层为10个神经元节点。

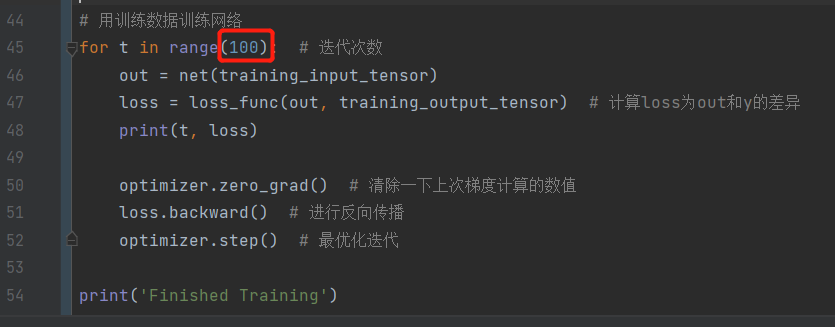


最后，设置优化器，设定学习速率（每次优化率）为0.02。选用损失函数BCEWithLogitsLoss（Sigmoid+BCELoss）

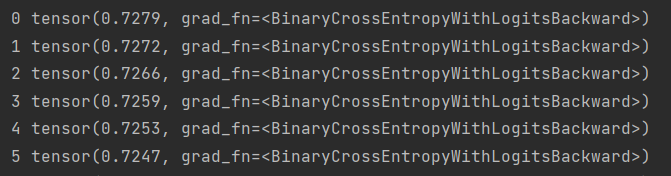


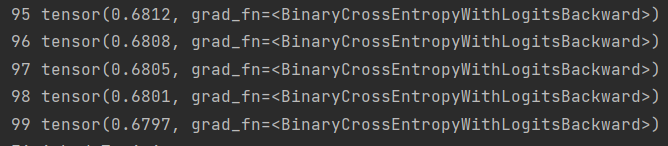
* 1. 开始训练

开始迭代训练，首先设定迭代次数为100，观察损失函数输出结果

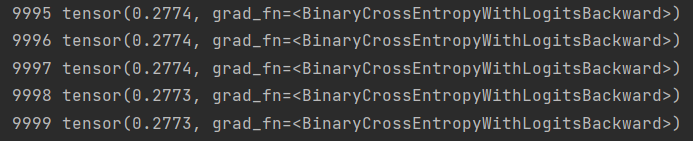


迭代训练100次后，损失函数结果从0.7279降为0.6797



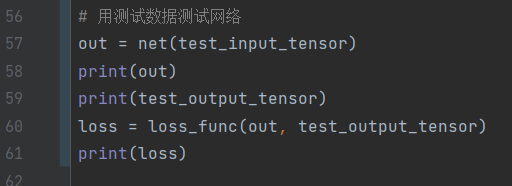


调整迭代为10000次，最终的损失函数结果为0.2773

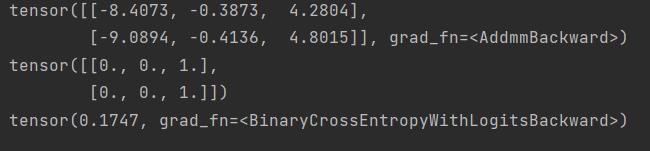


* 1. 测试模型结果

使用优化后的模型，输入测试input张量，将得到的输出张量与测试集比较：



得到损失函数结果为0.1747，证明根据模型得到的输出与测试数据偏差较小，比较满足预期。



* 1. 待优化

后续可尝试使用其他神经网络结构（如深层次神经网络mnist），并进行参数调整，使回归更加快速准确。

另外，中间的训练数据和最终测试结果的呈现，可引入matplotlib.pyplot进行多样化的报表来展现。

1. 总结

本文使用Pytorch中的线性回归网络，结合LSTM实现了对上市公司财务预警的简单建模分析，尝试使用深度学习框架解决现实中的分类问题。实际情况下公司的财务指标更加多样复杂，需要尝试引入更多指标，并需要多层神经网络、增大数据集达到优化训练的效果。

参考文献：

[python中用pytorch机器学习分类预测银行客户流失模型](http://tecdat.cn/python%E4%B8%AD%E7%94%A8pytorch%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0%E5%88%86%E7%B1%BB%E9%A2%84%E6%B5%8B%E9%93%B6%E8%A1%8C%E5%AE%A2%E6%88%B7%E6%B5%81%E5%A4%B1%E6%A8%A1%E5%9E%8B/)

[从Linear到LSTM，5种方法由浅入深预测股价](https://zhuanlan.zhihu.com/p/72992893)