**实验六 英文数字语音识别 作业报告**

1. 实验目的
   1. 掌握基于深度神经网络的语音识别基本原理
   2. 掌握基于深度神经网络的语音识别实现方法
2. 实验要求
3. 使用LibROSA包对语音信号进行特征提取。
4. 基于深度学习框架TensorFlow构建卷积神经网络模型，实现英文数字语音zero-nine的识别。
5. 实验原理

本实验主要是使用LibROSA语音库实现语音信号的特征提取与CNN模型的搭建，具体原理如下：

1. **特征提取**

主要提取MFCC（Mel Frequency Cepstral Coefficents）特征，MFCC特征是一种在自动语音识别和说话人识别（声纹识别）中广泛使用的特征。本实验是通过 LibROSA库函数实现了该特征提取。传统提取MFCC特征的过程如下：

1. 先对语音进行预加重、分帧和加窗；
2. 对每一个短时分析窗，通过FFT得到对应的频谱；
3. 将上面的频谱通过Mel滤波器组得到Mel频谱；
4. 在Mel频谱上面进行倒谱分析（取对数，做逆变换，实际逆变换一般是通过DCT离散余弦变换来实现，取DCT后的第2个到第13个系数作为MFCC系数），获得Mel频率倒谱系数MFCC，这个MFCC就是这帧语音的特征；

语音就可通过一系列的倒谱向量来表示，每个向量就是每帧的MFCC特征向量。通过这些倒谱向量就可对语音分类器进行训练和识别。

1. **卷积神经网络**

卷积神经网络主要是对上面的特征向量进行训练,实现数字的分类。CNN的输入是shape为[20,100]数据矩阵，分别经过四种size的卷积核（多种核可捕获更多特征）：[2,100]、[3,100]、[4,100]、[5,100]。四种卷积核的个数都是64个，分别得到64个长度为18的向量、64个长度为19的向量、64个长度为17的向量、64个长度为16的向量。然后取每个向量的最大值，取最大值的原因是保留每个卷积核捕获到的最大特征。把这些最大特征拼凑在一块作为x经过卷积后的特征向量(size=256)，后面接全连接层以及输出层。

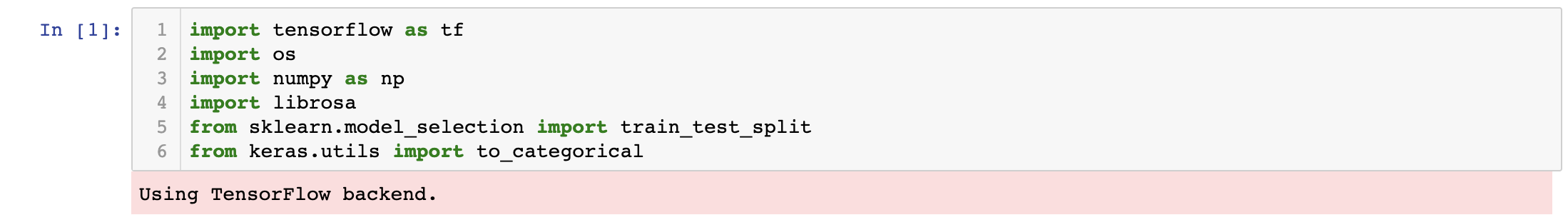
1. 实验所用工具及数据集
2. 主要工具

Python-3.5+、TensorFlow-1.3.0、librosa-0.6、Numpy-1.13.1

1. [数据集](https://pan.baidu.com/s/1N1DPirchGagsutwcODXuHA)（提取码: 9kiq）

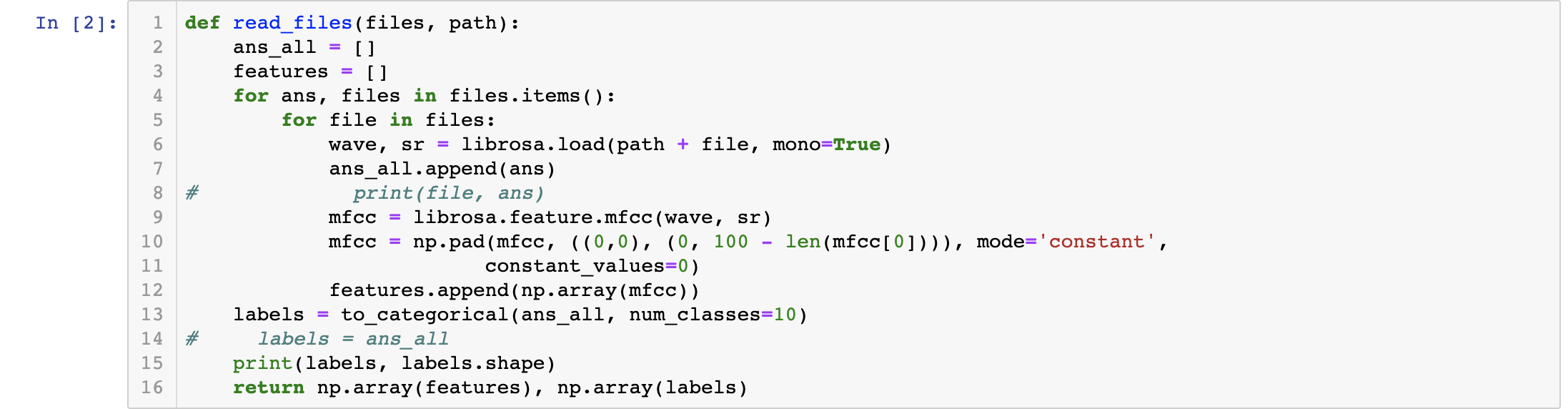
本实验使用英文数字语音数据集，共3800余条。其中训练集占70%，验证集20%，测试集10%。

1. 实验步骤与方法
2. **代码依赖**

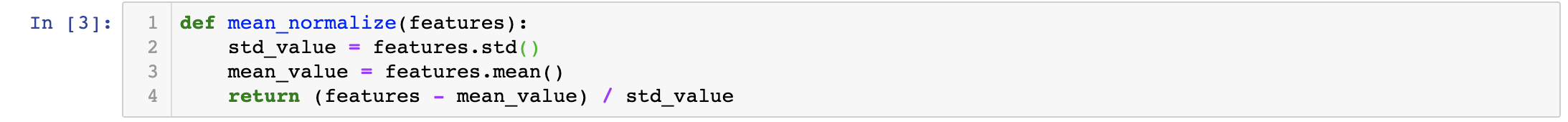


1. **数据预处理**

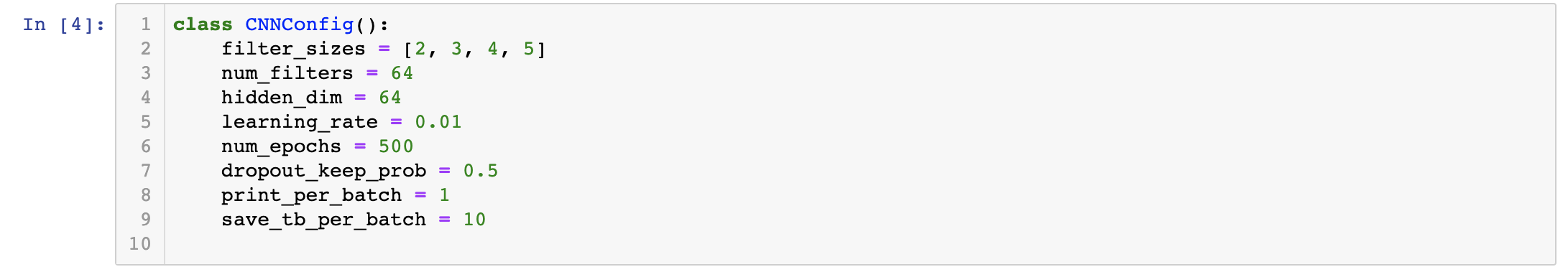
将语音数据转为mfcc特征



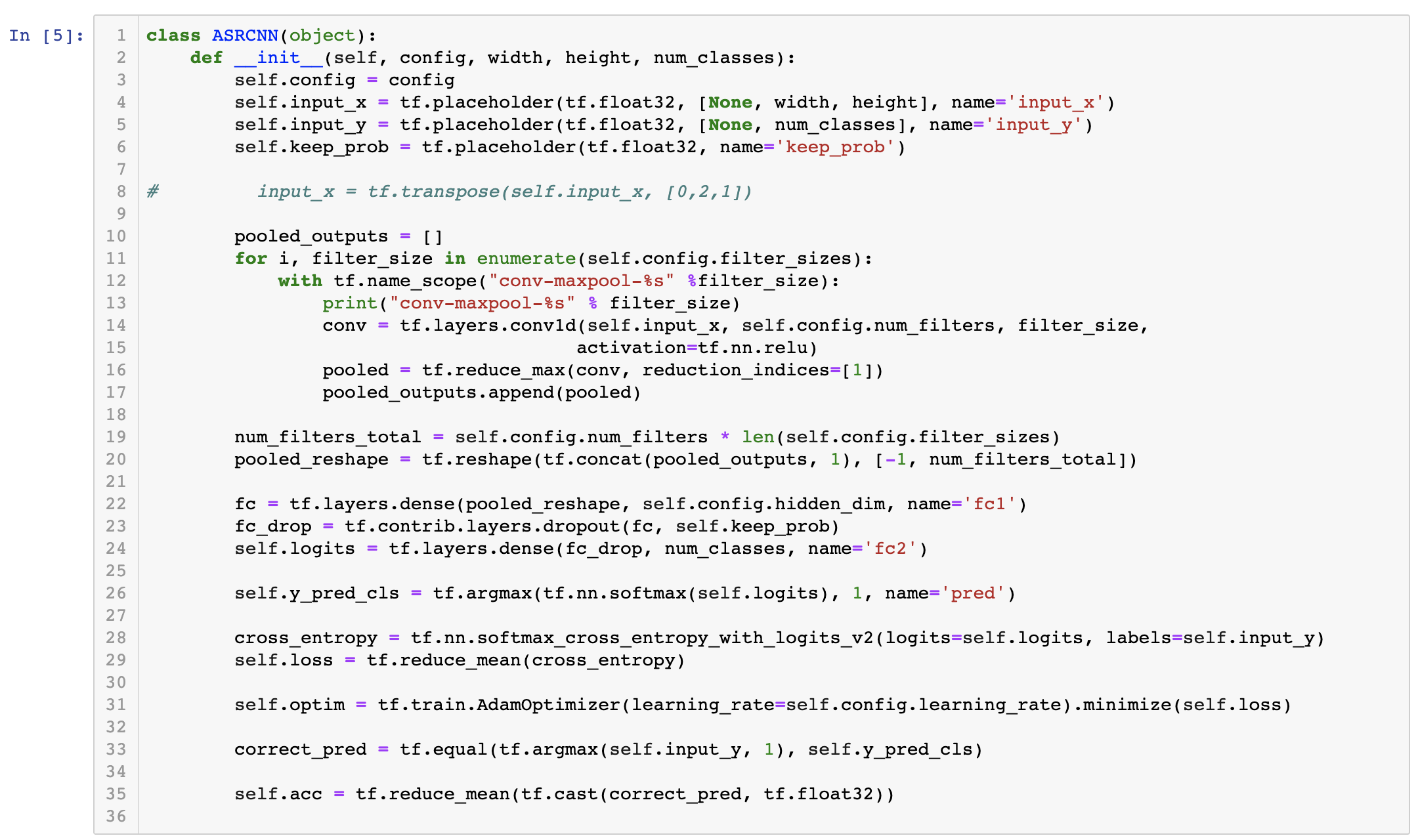
数据标准化



1. **模型参数**

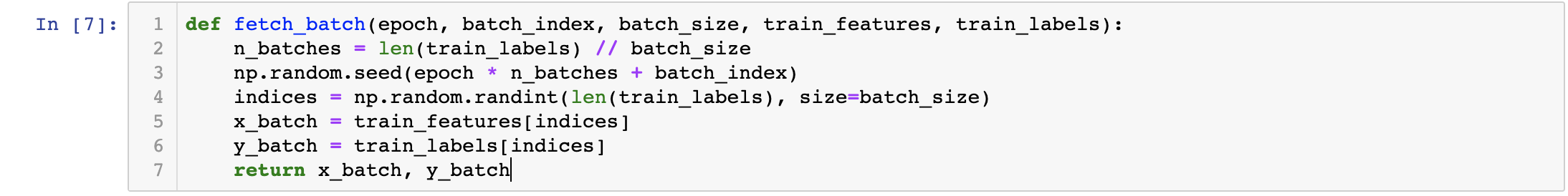


采用卷积加全连接搭建ASRCNN

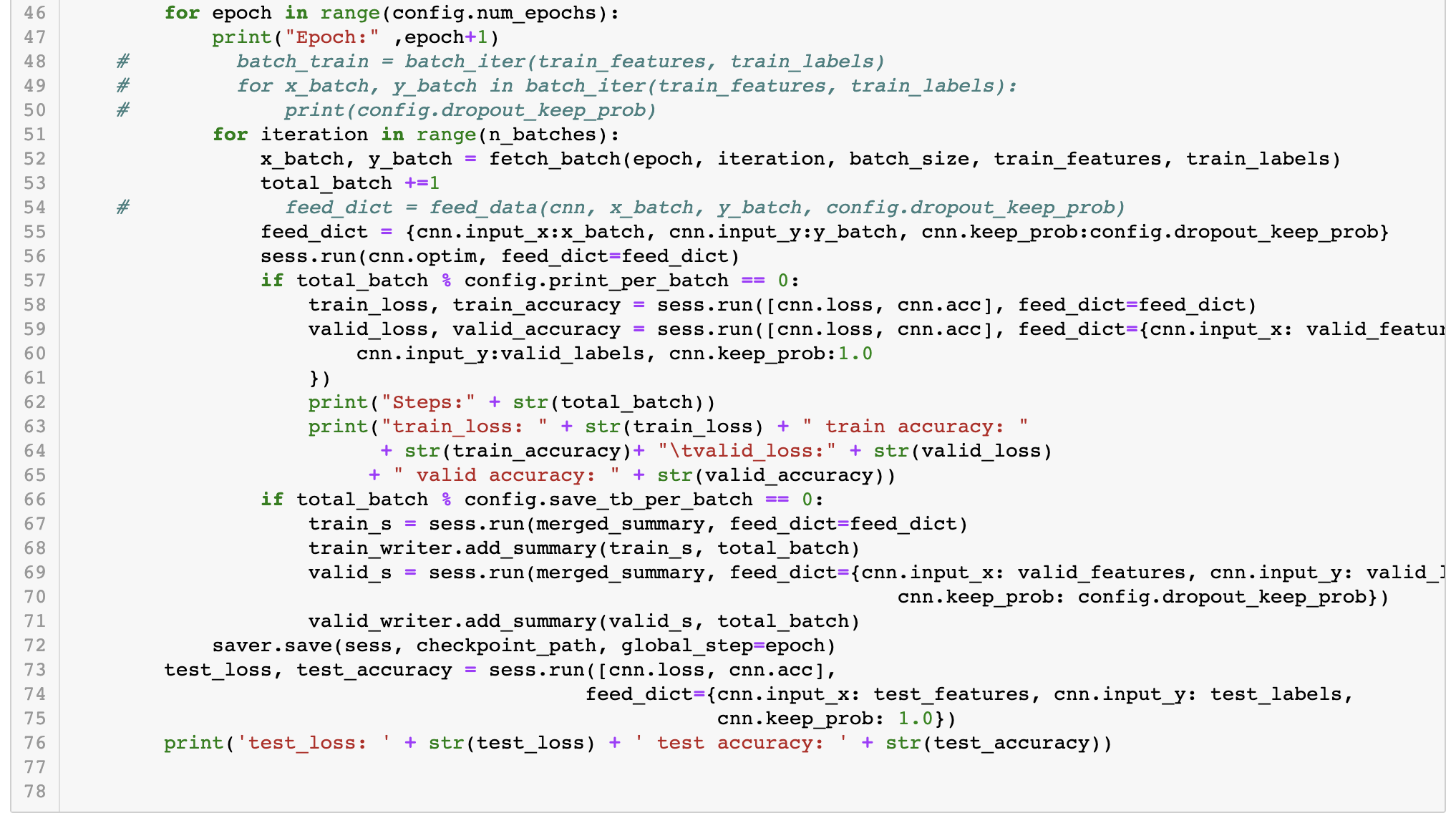


1. **模型训练**

批量读取数据训练







最后读取训练集，测试集，验证集进行模型训练，并输出训练结果：

