

**模式识别大作业**

题 目 MNIST手写数字识别

学 院 信息科学与工程

专 业 信息与通信工程

组 员 于俊杰

指导教师 赵海涛

**完成日期： 2019 年 11 月24日**

**模式识别作业报告**

组员：于俊杰

**一、MNIST手写数字识别简介**

MNIST（“国家标准技术研究院”）是计算机视觉的事实上的“ hello world”数据集。自从1999年发布以来，这个经典的手写图像数据集已成为基准分类算法的基础。随着新的机器学习技术的出现，MNIST仍然是研究人员和学习者的可靠资源。项目的目标是从数万个手写图像的数据集中正确地识别数字。

MNIST数据库由NIST的特殊数据库构成，它包含手写数字的二进制图像及其对应标签。 MNIST训练集由来自SD-3的30,000个样例和来自SD-1的30,000个样例组成。测试集由来自SD-3的5,000个图案和来自SD-1的5,000个图案组成。

**二、整体解决方案**

本项目将会采用多层感知机及CNN卷积神经网络两种方式进行识别和对比。

**2.1数据结构分析**

首先考察原始数据，数据来源于http://yann.lecun.com/exdb/mnist/，包含四个数据文件如图1

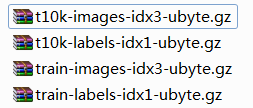


图 1 数据文件

按照从上到下的次序依次为：测试样本、测试标签、训练样本及训练标签。

样本图片为28x28像素，在数据集上中体现为长度是784的行向量，每个元素的值为0-255。

**2.2数据读入**

由于数据集是以不同压缩包的形式给出，因此需要现在程序中对其进行解压处理。这里直接使用了已有的函数模块。

然后需要将样本及标签匹配起来，并构造在程序中训练集和测试集。代码中通过构造Dataset类的方式来完成，不同于直接通过矩阵（或列表）的方式进行构造，用类的方式构造数据集使得数据的使用和划分更加明确。

需要注意的是，该类在初始化时将像素值0-255归一化到0-1之间，便于之后的数据计算和处理。

**2.3多层感知机模型建立及结果**

在获得了上述训练数据的基础上，搭建多层感知机模型。

多层感知机（MLP，Multilayer Perceptron）也叫人工神经网络（ANN，Artificial Neural Network），除了输入输出层，它中间可以有多个隐层，最简单的MLP只含一个隐层，即三层的结构，如图2。

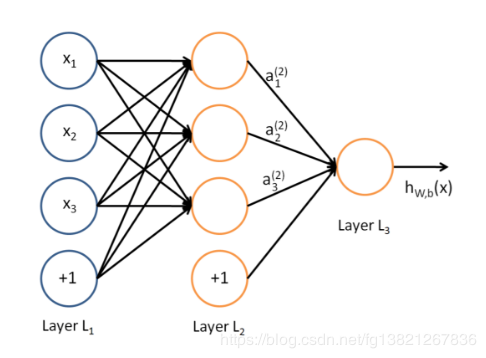


图 2 多层感知机模型

     图2可以看到，多层感知机层与层之间是全连接的。其最底层是输入层，中间是隐藏层，最后是输出层。

本项目中，搭建的网络使用两层隐层，具体参数如图3。

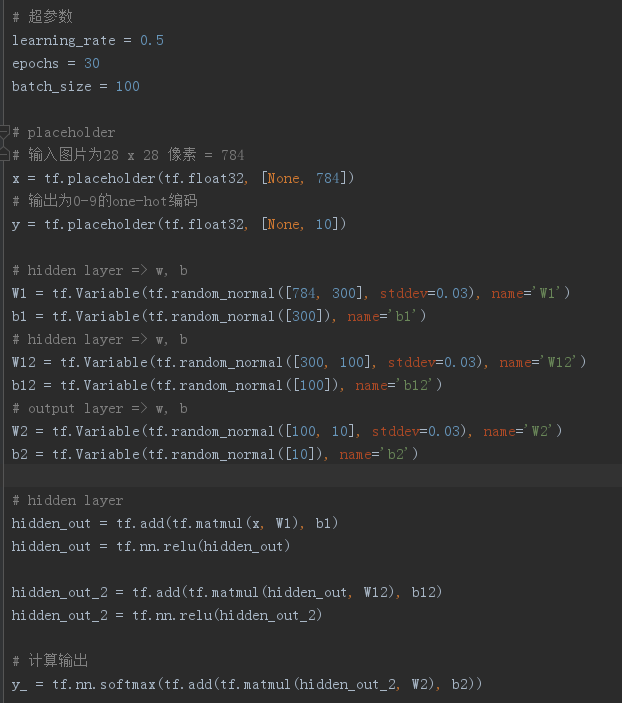


图 3 多层感知机参数

两个隐层的节点数量分别为300和100，而输出层节点为10，这是为了匹配10个不同输出的one-hot编码，也是分类任务的常规操作。

在经过30批次的训练（每批次100组），模型的准确度就已经达到了98%，如图4所示。

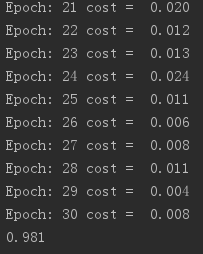


图 4 多层感知机识别结果

实验初始，仅使用了一层含300节点的隐层，准确率为96%，而增加隐层后，准确率提升了2%，而训练时间有小幅增加。这是因为作为图像的数据集本身具有高维特征，在增加了隐藏层后，这些高维特征被挖掘出来，为模型分类提供了参考。

**2.4卷积神经网络**

本项目同样采用了CNN卷积神经网络，具体上使用了Lenet网络。

手写字体识别模型LeNet5诞生于1994年，是最早的卷积神经网络之一。LeNet5通过巧妙的设计，利用卷积、参数共享、池化等操作提取特征，避免了大量的计算成本，最后再使用全连接神经网络进行分类识别，这个网络也是最近大量神经网络架构的起点。

LeNet5的网络结构示意图5所示：

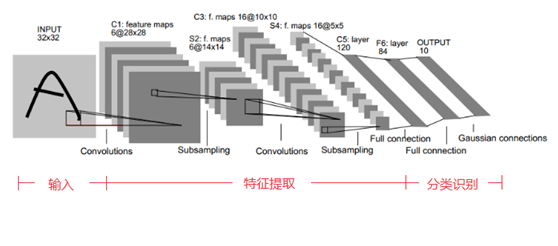


图 5 Lenet网络结构

从代码层面，图6展示了其具体参数和实现。

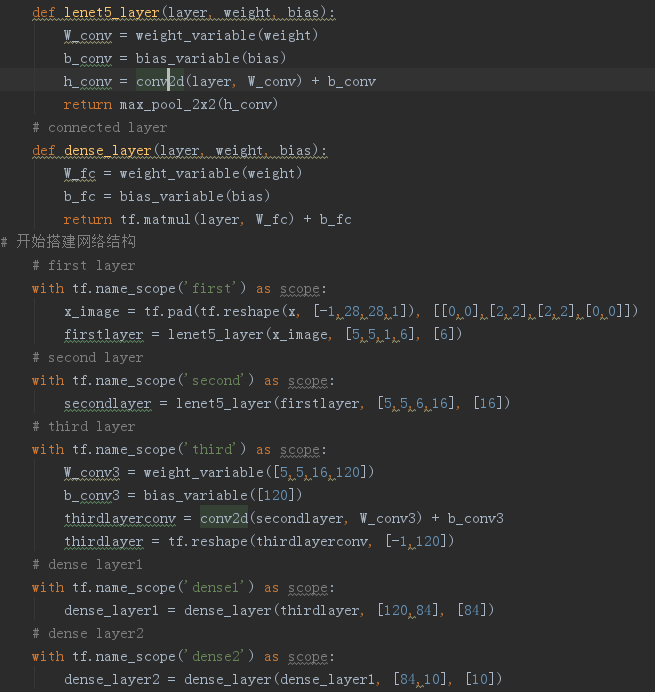


图 6 Lenet网络结构

从代码里容易看到，第一卷积层采用5x5的卷积核，输出6个特征图。第二、第三卷积层分别为16和120输出，其中均间隔了下采样层。

经过模型训练20000轮后，测试集准确率亦可以达到97.4%，如图7。

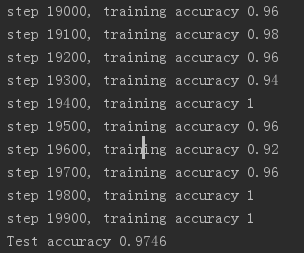


图 7 Lenet测试结果

**四、作业总结**

本次作业以MNIST手写数字识别为题，采用两种机器学习的方式来实现，均取得了较高的准确率，通过实现和分析代码，提高了自身的代码能力及对机器学习的了解程度。