

详解MNIST数据集

笔记本： 深度学习

创建时间： 2018/5/26 7:53

更新时间： 2018/5/26 8:09

作者： beyourselfwb@163.com

标签： 数据集

URL: https://blog.csdn.net/simple_the_best/article/details/75267863

MNIST 数据集可在 <http://yann.lecun.com/exdb/mnist/> 获取, 它包含了四个部分:

Training set images: train-images-idx3-ubyte.gz (9.9 MB, 解压后 47 MB, 包含 60,000 个样本)

Training set labels: train-labels-idx1-ubyte.gz (29 KB, 解压后 60 KB, 包含 60,000 个标签)

Test set images: t10k-images-idx3-ubyte.gz (1.6 MB, 解压后 7.8 MB, 包含 10,000 个样本)

Test set labels: t10k-labels-idx1-ubyte.gz (5KB, 解压后 10 KB, 包含 10,000 个标签)

MNIST 数据集来自美国国家标准与技术研究所, **National Institute of Standards and Technology (NIST)**.

每张图片由 $28 * 28$ 个像素点组成, 扁平化之后就是 784, 也就是input is (n, 784)

原始的label是一个单独的整数, 正常训练的时候如何转成one-hot vector呢?

```
In [9]: imgs, labels = dataiter.next()
```

```
In [11]: imgs.size()
```

```
Out[11]: torch.Size([64, 1, 28, 28])
```

```
In [12]: labels.size()
```

```
Out[12]: torch.Size([64])
```

```
In [14]: labels
```

```
Out[14]: tensor([ 0,  9,  1,  4,  5,  3,  2,  9,  0,  6,  8,  9,  3,  7,  
                  5,  6,  8,  1,  2,  6,  3,  4,  2,  8,  1,  6,  3,  5,  
                  9,  6,  0,  2,  2,  5,  9,  8,  1,  3,  9,  5,  2,  1,  
                  9,  1,  5,  6,  9,  1,  8,  4,  5,  4,  9,  0,  2,  3,  
                  2,  9,  6,  6,  1,  7,  9,  0])
```