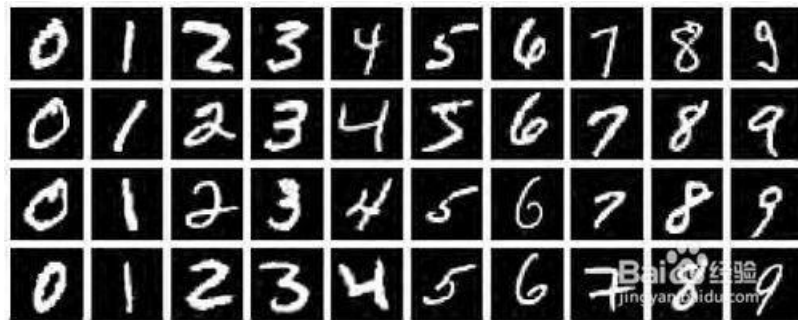




7.3 手写数字数据集

■ MNIST数据集 Mixed National Institute of standards and Technology database

- New York University, **Yann LeCun**
- **60000**条训练数据和**10000**条测试数据
- 由250个不同的人手写而成
- **28×28**像素, 灰度图像
- 存储在28×28的**二维数组**中



□ 下载MNIST数据集

```
import tensorflow as tf  
mnist = tf.keras.datasets.mnist  
(train_x, train_y), (test_x, test_y) = mnist.load_data()
```

```
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-datasets/mnist.npz  
11493376/11490434 [=====] - 18s 2us/step
```

下载路径: C:\Users\Administrator\.keras\datasets\mnist.npz

C:\Users\Lab\.keras\datasets\mnist.npz

当前Windows用户



□ 训练集和测试集的长度

```
print("Training set:",len(train_x))  
print("Testing set:",len(test_x))
```

运行结果:

```
Training set: 60000  
Testing set: 10000
```

□ 输出图像数据和标记数据的形状

```
print("train_x:",train_x.shape,train_x.dtype)  
print("train_y:",train_y.shape,train_y.dtype)
```

运行结果:

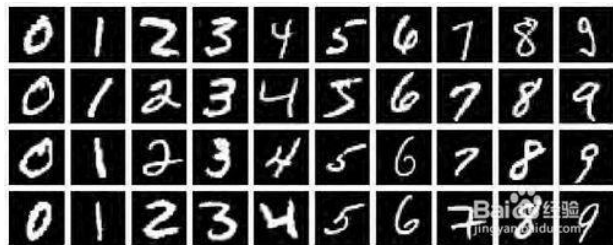
```
Train_x: (60000, 28, 28) : uint8  
Train_y: (60000,) uint8
```



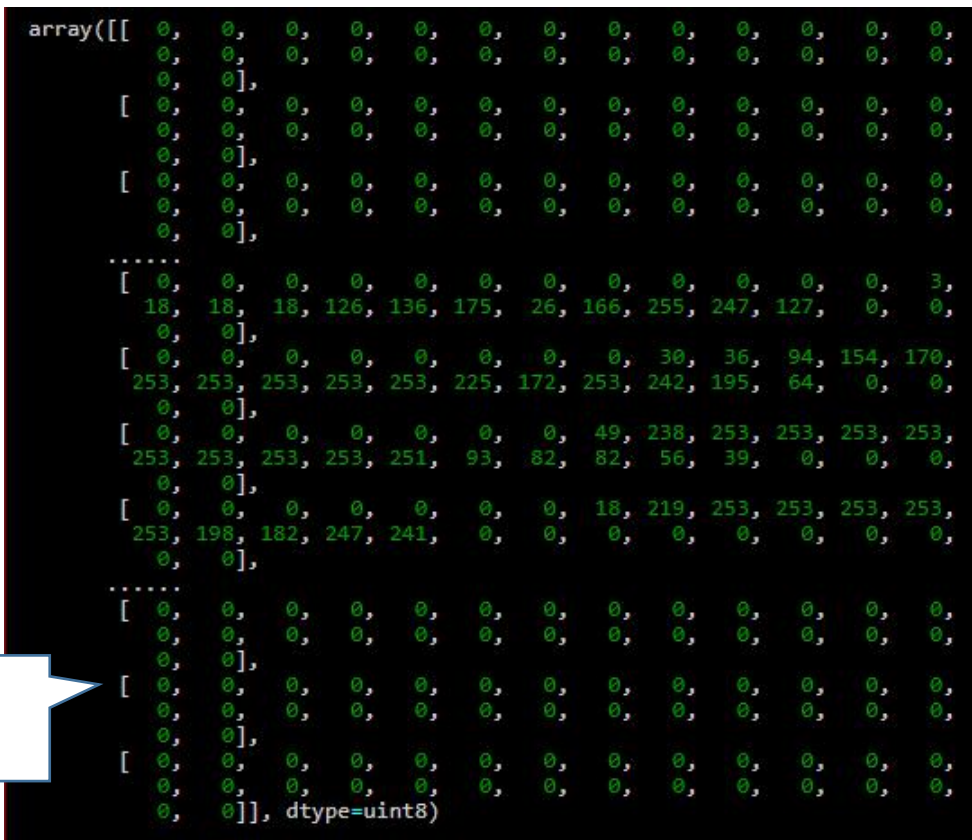
显示手写数字图片

- 输出数据集中的第1个样本

```
train_x[0]
```



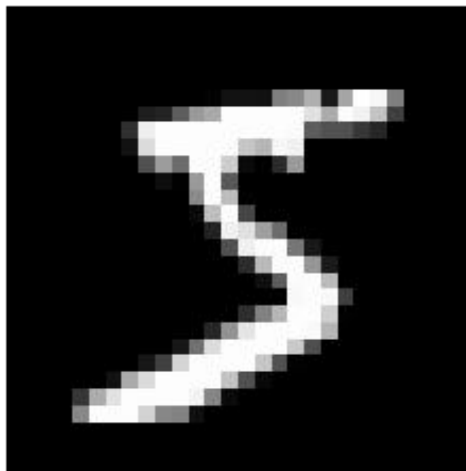
0-黑色; 255-白色
数值越大, 亮度越高



□ 显示图片

```
import matplotlib.pyplot as plt  
  
plt.axis("off")  
plt.imshow(train_x[0], cmap="gray")  
plt.show()
```

运行结果:



图片标记

```
>>> train_y[0]  
5
```



例：随机显示4幅数字图片

```
import tensorflow as tf
import numpy as np
import matplotlib.pyplot as plt

mnist = tf.keras.datasets.mnist
(train_x, train_y), (test_x, test_y) = mnist.load_data()

for i in range(4):
    num = np.random.randint(1, 50000)
    plt.subplot(1, 4, i+1)
    plt.axis("off")
    plt.imshow(train_x[num], cmap='gray')
    plt.title(train_y[num])

plt.show()
```

索引值是随机产生的，
每次运行结果都会不同

