

[Load the MNIST Dataset from Local Files](#)

[Overview](#)

[Example 1 Part 1 - Downloading the MNIST dataset](#)

[Example 1 Part 2 - Loading MNIST into NumPy Arrays](#)

[API](#)

Load the MNIST Dataset from Local Files

A utility function that loads the **MNIST** dataset from byte-form into NumPy arrays.

```
from mlxtend.data import loadlocal_mnist
```

Overview

The MNIST dataset was constructed from two datasets of the US National Institute of Standards and Technology (NIST). The training set consists of handwritten digits from 250 different people, 50 percent high school students, and 50 percent employees from the Census Bureau. Note that the test set contains handwritten digits from different people following the same split.

The MNIST dataset is publicly available at <http://yann.lecun.com/exdb/mnist/> and consists of the following four parts: - Training set images: train-images-idx3-ubyte.gz (9.9 MB, 47 MB unzipped, and 60,000 samples) - Training set labels: train-labels-idx1-ubyte.gz (29 KB, 60 KB unzipped, and 60,000 labels) - Test set images: t10k-images-idx3-ubyte.gz

(1.6 MB, 7.8 MB, unzipped and 10,000 samples) - Test set labels: t10k-labels-idx1-ubyte.gz (5 KB, 10 KB unzipped, and 10,000 labels)

Features

Each feature vector (row in the feature matrix) consists of 784 pixels (intensities) -- unrolled from the original 28x28 pixels images.

- Number of samples: 50000 images
- Target variable (discrete): Uniformly distributed class labels 0-9 corresponding to the respective handwritten digit shown in the image.

References

- Source: <http://yann.lecun.com/exdb/mnist/>
- Y. LeCun and C. Cortes. Mnist handwritten digit database. AT&T Labs [Online]. Available: <http://yann.lecun.com/exdb/mnist>, 2010.

Example 1 Part 1 - Downloading the MNIST dataset

1) Download the MNIST files from Y. LeCun's website

- <http://yann.lecun.com/exdb/mnist/train-images-idx3-ubyte.gz>
- <http://yann.lecun.com/exdb/mnist/train-labels-idx1-ubyte.gz>
- <http://yann.lecun.com/exdb/mnist/t10k-images-idx3-ubyte.gz>
- <http://yann.lecun.com/exdb/mnist/t10k-labels-idx1-ubyte.gz>

for example, via

```
curl -O http://yann.lecun.com/exdb/mnist/train-images-idx3-ubyte.gz
```

2) Unzip the downloaded gzip archives

for example, via

```
gunzip t*-ubyte.gz
```

Example 1 Part 2 - Loading MNIST into NumPy Arrays

```
from mlxtend.data import loadlocal_mnist
```

```
X, y = loadlocal_mnist(
    images_path='/Users/Sebastian/Desktop/train-images-idx3-ubyte',
    labels_path='/Users/Sebastian/Desktop/train-labels-idx1-ubyte')
```

```
print('Dimensions: %s x %s' % (X.shape[0], X.shape[1]))
print('\n1st row', X[0])
```

Dimensions: 60000 x 784

```
1st row [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0  0  3 18 18 18 126 136 175 26 166 255
247 127  0  0  0  0  0  0  0  0  0  0  0  0  30 36 94 154
170 253 253 253 253 253 225 172 253 242 195 64  0  0  0  0  0  0
 0  0  0  0  0 49 238 253 253 253 253 253 253 253 251 93 82
82 56 39  0  0  0  0  0  0  0  0  0  0  0  0 18 219 253
253 253 253 253 198 182 247 241  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0  0 80 156 107 253 253 205 11  0 43 154
```

```
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 14 1 154 253 90 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 139 253 190 2 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 11 190 253 70 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 35 241
225 160 108 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 81 240 253 253 119 25 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 45 186 253 253 150 27 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 16 93 252 253 187
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 249 253 249 64 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 46 130 183 253
253 207 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 39 148 229 253 253 253 250 182 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 24 114 221 253 253 253
253 201 78 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 23 66 213 253 253 253 253 198 81 2 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 18 171 219 253 253 253 253 195
80 9 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
55 172 226 253 253 253 253 244 133 11 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 136 253 253 253 212 135 132 16
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```
import numpy as np

print('Digits: 0 1 2 3 4 5 6 7 8 9')
print('labels: %s' % np.unique(y))
print('Class distribution: %s' % np.bincount(y))
```

```
Digits: 0 1 2 3 4 5 6 7 8 9
labels: [0 1 2 3 4 5 6 7 8 9]
Class distribution: [5923 6742 5958 6131 5842 5421 5918 6265 5851 5949]
```

Store as CSV Files

```
np.savetxt(fname='/Users/Sebastian/Desktop/images.csv',
           X=X, delimiter=',', fmt='%d')
np.savetxt(fname='/Users/Sebastian/Desktop/labels.csv',
           X=y, delimiter=',', fmt='%d')
```

API

`loadlocal_mnist(images_path, labels_path)`

Read MNIST from ubyte files.

Parameters

- `images_path` : str
path to the test or train MNIST ubyte file
- `labels_path` : str
path to the test or train MNIST class labels file

Returns

- `images` : [n_samples, n_pixels] numpy.array
Pixel values of the images.
- `labels` : [n_samples] numpy array
Target class labels

Examples

For usage examples, please see

http://rasbt.github.io/mlxtend/user_guide/data/loadlocal_mnist/

