

MNIST Example

Carlos Carvalho, Mladen Kolar, Rob McCulloch

10/21/2015

In this example, we will explore the famous MNIST handwritten digits data set.

Data are provided by Yann LeCun and can be downloaded here: <http://yann.lecun.com/exdb/mnist/index.html>.

For convenience, the data can be also downloaded from a GitHub [repository](#). For example, you can clone the repository using the following command:

```
git clone https://github.com/ChicagoBoothML/DATA\_\_\_\_LeCun\_\_\_\_MNISTDigits.git
```

Remark: Data can also be downloaded from Kaggle: <https://www.kaggle.com/c/digit-recognizer/data>. Note that the data available from the Kaggle's website is not partitioned in the same way into training and test sets. Below, I will be using data from Yann LeCun's website.

The MNIST data set has a training set of 60,000 examples, and a test set of 10,000 examples. The digits have been size-normalized and centered in a fixed-size image.

Each observation is a grey-scale image sized 28 by 28 pixels. The columns are the pixel numbers, ranging from pixel 0 to pixel 783 (784 total pixels), which have elements taking values from 0 to 255 (white is 0 and 255 is black). Thus, our observations each have 784 feature values.

The goal is to build a model that will be presented with an image of a numerical digit (0-9) and the model must predict which digit is being shown.