



InClass Prediction Competition

# MNIST Digit Classification I

## Classification of handwritten digits.

[Overview](#)[Data](#)[Kernels](#)[Discussion](#)[Leaderboard](#)[Rules](#)

Overview

[Description](#)

## Description

Evaluation


In this exercise you will use k-Nearest Neighbors (k-NN), manifold tangent distances, and related ideas we discussed in lecture to classifying the handwritten digits from the database MNIST (Modified National Institute of Standards). We will also revisit this data set using other machine learning methods later in the course. The input handwritten digits are 28x28 grey-scale images (784 pixels). The task is to use the training set to develop a classifier able to correctly predict which digit is shown in each 28x28 image.

In the k-NN approach, the distance metric captures our notion of similarity and is central to the performance of the classifier. Along the lines we discussed in lecture, you are to investigate distance metrics and related methods that incorporate prior information about likely intra-class invariances that could be helpful in enhancing classification. You can be creative in how you build on the general mathematical frameworks we developed in lecture for using transformations and invariances. For instance, we showed techniques in lecture that capture invariances using associated geometric manifolds in image space. The goal is to classify the image as accurately as possible into the correct digit class. Good luck on this task. Look forward to seeing your results.

This data is shared under the [Creative Commons Attribution-Share Alike 3.0 license](#). For more information on the history of the MNIST digit classification problem and past performance of various machine learning methods on this task, there is a nice discussion on Yann LeCun's webpage [here](#).

## Author of Competition

This competition was setup for a data science course by Paul J. Atzberger, <http://atzberger.org/>

 InClass Prediction Competition

# MNIST Digit Classification I

## Classification of handwritten digits.

Overview **Data** Kernels Discussion Leaderboard Rules

### Data Description

This is the MNIST (Modified National Institute of Standards) database of handwritten digits. The images are grey-scale 28x28 images. They are arranged in a .csv spreadsheet with the first column the label of the digit and the next 784 columns the pixel values of the image.

### File descriptions

- train\_small.csv - a training subset of the 10,000 images (helpful when prototyping so algorithms run faster).
- train\_full.csv - the full training set of 60,000 images (should be used for your full production level training).
- test\_medium.csv - the test set (predict labels for this data set).
- test\_medium\_sample.csv - a sample submission file in the format required for scoring.

### Data fields

- label - label associated with the image of the digit.
- pixel\_XXX - one of the 784 pixel values of the 28x28 image.

Data (18 MB)

API

★ kaggle competitions download -c atzb-mnist-digit-t...

🔍

Download All



#### Data Sources

- ↳ test\_all.csv 10000 x 785
- ↳ test\_mediu... 1000 x 785
- ↳ test\_mediu... 1000 x 786
- ↳ test\_small.csv 100 x 785
- ↳ train\_full.csv 60.0k x 786

#### About this file

No description yet

#### Columns

- ④ id
- label
- pixel\_000
- pixel\_001
- pixel\_002