Kaggle competition

https://www.kaggle.com/c/digit-recognizer/data

Assignment

1. Upload the solution (.csv file with labels, as described on the webpage) to the Kaggle competition Digit recognizer challenge.

2. Upload the code you used to generate the solution to OLAT (or send it to us)

Deadline: Dec 8th. 10:00 am

Getting Started Prediction Competition

Digit Recognizer

Learn computer vision fundamentals with the famous MNIST data

k   Kaggle · 2,537 teams · Ongoing

Overview   Data   Notebooks   Discussion   Leaderboard   Rules   **Join Competition**

Overview

Description

Tutorial

Start here if...

You have some experience with R or Python and machine learning basics, but you're new to computer

The data files train.csv and test.csv contain gray-scale images of hand-drawn digits, from zero through nine.

Each image is 28 pixels in height and 28 pixels in width, for a total of 784 pixels in total. Each pixel has a single pixel-value associated with it, indicating the lightness or darkness of that pixel, with higher numbers meaning darker. This pixel-value is an integer between 0 and 255, inclusive.

The training data set, (train.csv), has 785 columns. The first column, called "label", is the digit that was drawn by the user. The rest of the columns contain the pixel-values of the associated image.

Each pixel column in the training set has a name like pixelx, where x is an integer between 0 and 783, inclusive. To locate this pixel on the image, suppose that we have decomposed x as x = i * 28 + j, where i and j are integers between 0 and 27, inclusive. Then pixelx is located on row i and column j of a 28 x 28 matrix, (indexing by zero).

For example, pixel31 indicates the pixel that is in the fourth column from the left, and the second row from the top, as in the ascii-diagram below.

Visually, if we omit the "pixel" prefix, the pixels make up the image like this:

```
000 001 002 003 ... 026 027
028 029 030 031 ... 054 055
056 057 058 059 ... 082 083
 |   |   |   |  ...  |   |
728 729 730 731 ... 754 755
756 757 758 759 ... 782 783
```

The test data set, (test.csv), is the same as the training set, except that it does not contain the "label" column.

Your submission file should be in the following format: For each of the 28000 images in the test set, output a single line containing the ImageId and the digit you predict. For example, if you predict that the first image is of a 3, the second image is of a 7, and the third image is of a 8, then your submission file would look like:

```
ImageId,Label
1,3
2,7
3,8
(27997 more lines)
```

The evaluation metric for this contest is the categorization accuracy, or the proportion of test images that are correctly classified. For example, a categorization accuracy of 0.97 indicates that you have correctly classified all but 3% of the images.

Data (15 MB)

API `kaggle competitions download -c digit-recognizer` ? ⬇ **Download All**

Data Sources

⊞ sample_submission....	28.0k x 2
⊞ test.csv	28.0k x 784
⊞ train.csv	42.0k x 785

About this file

The file contains the test data, train data and a sample submission format.

Columns

\# ImageId datos

\# Label Predictions