Submission
✔ **Ran successfully**
Submitted by Marsh 10 months ago

Public Score
0.96757

Sklearn - LightGBM

Python notebook using data from Digit Recognizer · 4,302 views · 10mo ago

I am experimenting with simple and fast solutions. This LightGBM model scored above 0.96 on the public leaderboard. It took approx. 2 minutes to train on a cpu. I did not do any parameter tuning.

It's nice to discover that it's possible to get such a high accuracy with so little work.

In [1]:
```
import pandas as pd
```

In [2]:
```
# read the data into a pandas datafrome
df_train = pd.read_csv('../input/train.csv')
df_test = pd.read_csv('../input/test.csv')

print(df_train.shape)
print(df_test.shape)
```

```
(42000, 785)
(28000, 784)
```

Define X and y

In [3]:
```
X = df_train.drop('label', axis=1)
y = df_train['label']

X_test = df_test

print(X.shape)
print(y.shape)
print(X_test.shape)
```



```
(28000, 784)
```

LightGBM Classifier

In [4]:
```
%%time

from lightgbm import LGBMClassifier

lgbm = LGBMClassifier(objective='multiclass', random_state=5)
```

```
lgbm.fit(X, y)

y_pred = lgbm.predict(X_test)
```

```
CPU times: user 4min 24s, sys: 1.92 s, total: 4min 26s
Wall time: 2min 14s
```

In [5]:

```
y_pred.shape
```

Out[5]:

```
(28000,)
```

Create a submission file

In [6]:

```
# The index should start from 1 instead of 0
df = pd.Series(range(1,28001),name = "ImageId")

ID = df

submission = pd.DataFrame({'ImageId':ID,
                           'Label':y_pred,
                          }).set_index('ImageId')

submission.to_csv('mnist_lgbm.csv', columns=['Label'])
```

In [7]:

```
submission.head()
```

Out[7]:

	Label
ImageId	
1	2
2	0
3	9
4	9
5	3

Thank you for reading.

In [ ]:

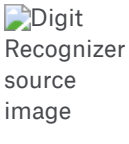
Did you find this Kernel useful?
Show your appreciation with an upvote

4

Data

Data Sources

- Digit Recognizer
  - sample_submission.csv 28.0k x 2
  - test.csv 28.0k x 784
  - train.csv 42.0k x 785



Digit Recognizer

Learn computer vision fundamentals with the famous MNIST data

Last Updated: 7 years ago

About this Competition

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
```

Output Files

New Dataset | New Notebook | Download All

Output Files

- mnist_lgbm.csv

About this file

This file was created from a Kernel, it does not have a description.

mnist_lgbm.csv

1	ImageId	Label
2	1	2
3	2	0
4	3	9
5	4	9
6	5	3
7	6	7
8	7	0
9	8	3
10	9	0
11	10	3
12	11	5
13	12	7
14	13	4
15	14	0
16	15	4
17	16	3
18	17	3
19	18	1
20	19	9
21	20	0
22	21	9
23	22	1
24	23	1
25	24	5
26	25	7
27	26	4
28	27	2
29	28	7
30	29	4
31	30	7
32	31	7

Comments (0)

Click here to comment...