CSE 4/574 Project 1 (part 2)

Classification of Handwritten Digits

CSE4/574 Machine Learning
TA: Yu Liu
yl73@buffalo.edu

Project Tasks!

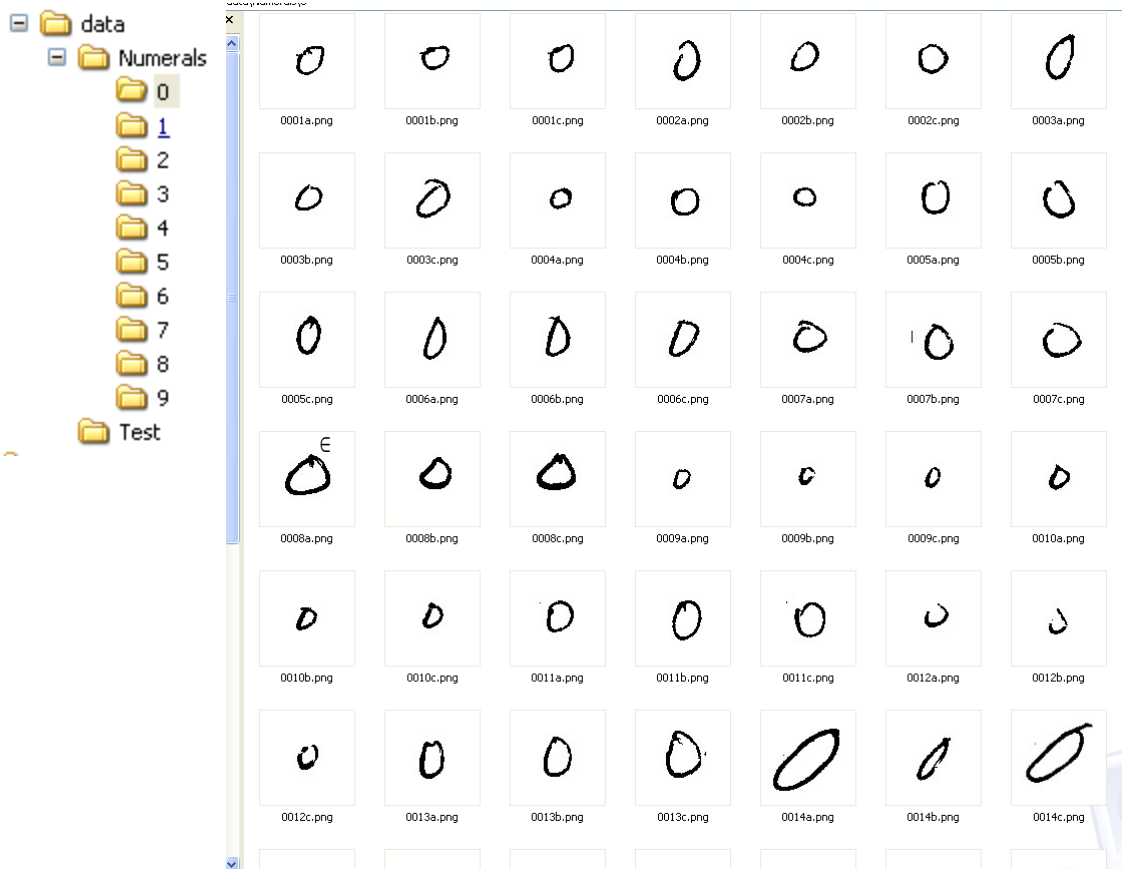
1. Process the data
2. Implement following classifiers

    Logistic Regression (LR)
    Neural Networks (NB)

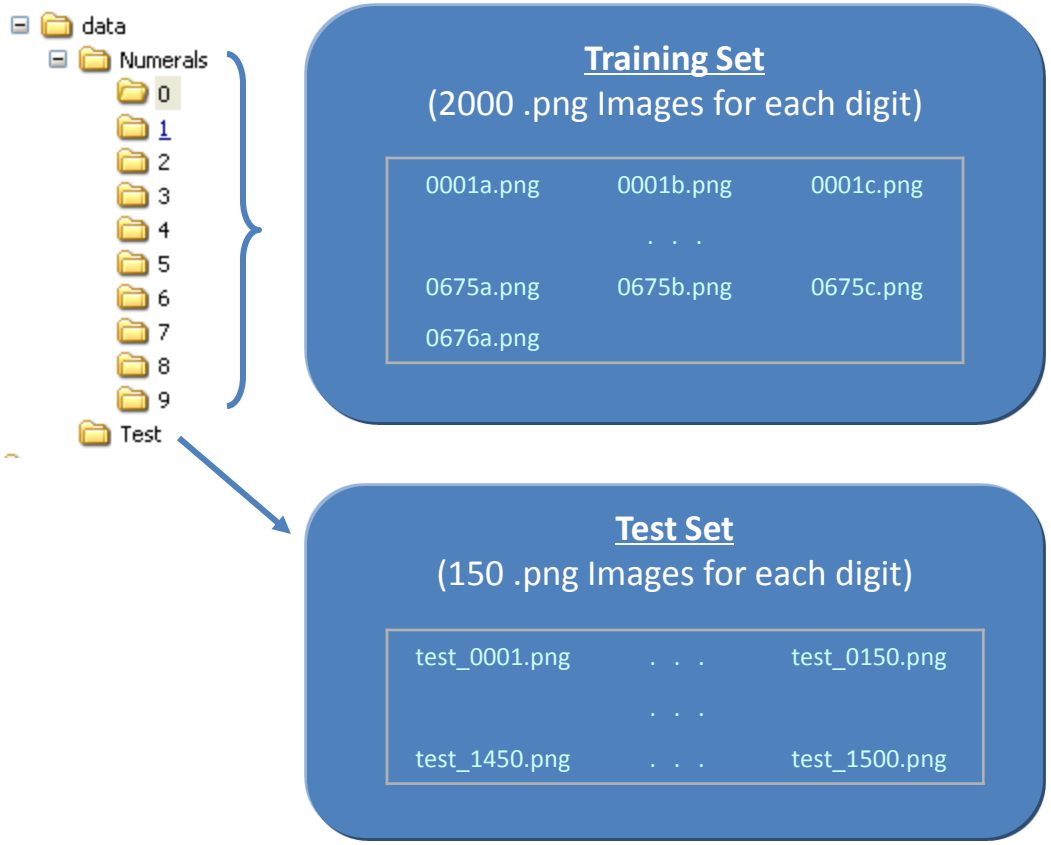
3. Evaluation and compare performance

Handwriting Digit Images

Sample images from the dataset:

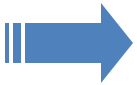


The Data Set

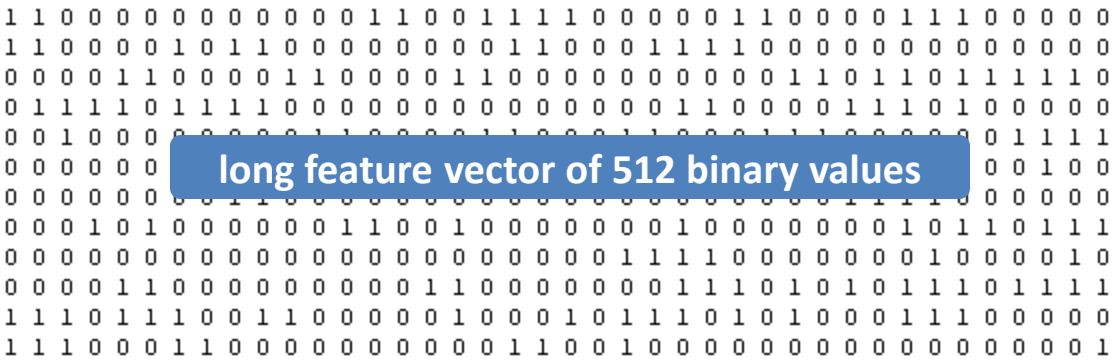

data
Numerals
0
1
2
3
4
5
6
7
8
9
Test
Training Set
(2000 .png Images for each digit)
0001a.png
0001b.png
0001c.png
. . .
0675a.png
0675b.png
0675c.png
0676a.png
Test Set
(150 .png Images for each digit)
test_0001.png
. . .
test_0150.png
. . .
test_1450.png
. . .
test_1500.png

Extract Features

filename.png



output.txt



- GSC features:
  - NO need to download images or the feature extractor!
  - Extracted features on **UBLearns**

- You are also encouraged to define your own feature…

Import Features in Matlab

- Matlab function:

  File -> Import Data …

  or

  >> load('0.txt');

The imported data is already a matrix containing features vectors for digit '0's.

Name	Value	Class
x0	<2000x512 double>	double

For the entire training data (all digits): 20000 x 512 matrix

Preparing Data Sets