

1. Title of Database: Optical Recognition of Handwritten Digits

2. Source:

E. Alpaydin, C. Kaynak
Department of Computer Engineering
Bogazici University, 80815 Istanbul Turkey
alpaydin@boun.edu.tr
July 1998

3. Past Usage:

C. Kaynak (1995) Methods of Combining Multiple Classifiers and Their Applications to Handwritten Digit Recognition, MSc Thesis, Institute of Graduate Studies in Science and Engineering, Bogazici University.

E. Alpaydin, C. Kaynak (1998) Cascading Classifiers, Kybernetika, to appear. <ftp://ftp.icsi.berkeley.edu/pub/ai/ethem/kyb.ps.Z>

4. Relevant Information:

We used preprocessing programs made available by NIST to extract normalized bitmaps of handwritten digits from a preprinted form. From a total of 43 people, 30 contributed to the training set and different 13 to the test set. 32x32 bitmaps are divided into nonoverlapping blocks of 4x4 and the number of on pixels are counted in each block. This generates an input matrix of 8x8 where each element is an integer in the range 0..16. This reduces dimensionality and gives invariance to small distortions.

For info on NIST preprocessing routines, see
M. D. Garriss, J. L. Blue, G. T. Candela, D. L. Dimmick, J. Geist, P. J. Grother, S. A. Janet, and C. L. Wilson, NIST Form-Based Handprint Recognition System, NISTIR 5469, 1994.

5. Number of Instances

optdigits.tra	Training	3823
optdigits.tes	Testing	1797

The way we used the dataset was to use half of training for actual training, one-fourth for validation and one-fourth for writer-dependent testing. The test set was used for writer-independent testing and is the actual quality measure.

6. Number of Attributes

64 input+1 class attribute

7. For Each Attribute:

All input attributes are integers in the range 0..16.
The last attribute is the class code 0..9

8. Missing Attribute Values

None

9. Class Distribution

Class:	No of examples in training set
0:	376
1:	389
2:	380
3:	389
4:	387
5:	376
6:	377
7:	387
8:	380

9: 382

Class: No of examples in testing set

0: 178

1: 182

2: 177

3: 183

4: 181

5: 182

6: 181

7: 179

8: 174

9: 180

Accuracy on the testing set with k-nn
using Euclidean distance as the metric

k = 1 : 98.00

k = 2 : 97.38

k = 3 : 97.83

k = 4 : 97.61

k = 5 : 97.89

k = 6 : 97.77

k = 7 : 97.66

k = 8 : 97.66

k = 9 : 97.72

k = 10 : 97.55

k = 11 : 97.89