1. **Optical Recognition of Handwritten Digits**

1. 数据库网址

http://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of+Handwritten+Digits

2. 数据库描述

【1.[数据集名称]数据集由[机构名或人名]采集；】The data used in our experiments were collected by E. Alpaydin, C. Kaynak, from Department of Computer Engineering, Bogazici University at July,1998.【2.用于[什么实验目的]】We used preprocessing programs made available by NIST to extract normalized bitmaps of handwritten digits from a preprinted form.【3】From a total of 43 people, 30 contributed to the training set and different 13 to the test set. 32x32 bitmaps are divided into non-overlapping blocks of 4x4 and the number of pixels is counted in each block. This generates an input matrix of 8x8 where each element is an integer in the range 0-16. Number of Attributes is 64 input+1 class attribute, All input attributes are integers in the range 0-16.The last attribute is the class code 0-9. This reduces dimensionality and gives invariance to small distortions. 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.【4】The database has 5620 samples, respectively belong to training with 3823 samples and testing with 1797 samples. The categories can be as as shown in Table 1.

Table 1 Category Distribution of handwritten dataset

|  |  |  |  |
| --- | --- | --- | --- |
| Categories | training | testing | Total Number of Samples |
| 0 | 376 | 178 | 554 |
| 1 | 389 | 182 | 571 |
| 2 | 380 | 177 | 557 |
| 3 | 389 | 183 | 572 |
| 4 | 387 | 181 | 568 |
| 5 | 376 | 182 | 558 |
| 6 | 377 | 181 | 558 |
| 7 | 387 | 179 | 566 |
| 8 | 380 | 174 | 554 |
| 9 | 382 | 180 | 562 |
| Total number of samples in total | 3823 | 1797 | 5620 |

3. 精简描述

The Data in our experiment were collected by E. Alpaydin, C. Kaynak, from Department of Computer Engineering, Bogazici University at July,1998. The dataset includes have 5620 samples, which used to preprocessing programs made available by NIST to extract normalized bitmaps of handwritten digits from a preprinted form. Through which, we divided the dataset into two part, training data set with 3823 samples and forecasting data set with 1797 samples.