**72.LED Display Domain**

1. 数据库网址

http://archive.ics.uci.edu/ml/datasets/LED+Display+Domain

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 Breiman,L., Friedman,J.H., Olshen,R.A., & Stone,C.J. (1984). from the Classification and Regression Trees.【2.用于[什么实验目的]】We used to classify the trees book.【3】This simple domain contains 7 Boolean attributes and 10 concepts, the set of decimal digits. Recall that LED displays contain 7 light-emitting diodes -- hence the reason for 7 attributes. The problem would be easy if not for the introduction of noise. In this case, each attribute value has the 10% probability of having its value inverted. All attribute values are either 0 or 1, according to whether the corresponding light is on or not for the decimal digit. Each attribute (excluding the class attribute, which is an integer ranging between 0 and 9 inclusive) has a 10% percent chance of being inverted. Each concept (digit) has the same theoretical probability distribution. The program randomly selects the attribute.【4】The database has 5620 samples, respectively belong to training with 3823 samples and testing with 1797 samples. The categories of LED dataset include ten categories, as shown in Table 1.

Table 1 Category Distribution of LED dataset [根据数据库绘制]

|  |  |  |  |
| --- | --- | --- | --- |
| Categories | Training | testing | Total Number of Samples |
| 0 | 200 | 50 | 250 |
| 1 | 200 | 50 | 250 |
| 2 | 200 | 50 | 250 |
| 3 | 200 | 50 | 250 |
| 4 | 200 | 50 | 250 |
| 5 | 200 | 50 | 250 |
| 6 | 200 | 50 | 250 |
| 7 | 200 | 50 | 250 |
| 8 | 200 | 50 | 250 |
| 9 | 200 | 50 | 250 |
| Total number of samples in total | 2000 | 500 | 2500 |

3. 精简描述

The Data in our experiment were collected by Breiman,L., Friedman,J.H., Olshen,R.A., & Stone,C.J. The dataset includes have 2500 samples, which used to classify the trees book. Through which, we divided the dataset into two part, training data set with 2000 samples and forecasting data set with 500 samples.