1. Title: Letter Image Recognition Data

2. Source Information
-- Creator: David J. Slate
-- Odesta Corporation; 1890 Maple Ave; Suite 115; Evanston, IL 60201
-- Donor: David J. Slate (dave@math.nwu.edu) (708) 491-3867
-- Date: January, 1991

3. Past Usage:
-- P. W. Frey and D. J. Slate (Machine Learning Vol 6 #2 March 91):
"Letter Recognition Using Holland-style Adaptive Classifiers".

The research for this article investigated the ability of several variations of Holland-style adaptive classifier systems to learn to correctly guess the letter categories associated with vectors of 16 integer attributes extracted from raster scan images of the letters. The best accuracy obtained was a little over 80%. It would be interesting to see how well other methods do with the same data.

4. Relevant Information:

The objective is to identify each of a large number of black-and-white rectangular pixel displays as one of the 26 capital letters in the English alphabet. The character images were based on 20 different fonts and each letter within these 20 fonts was randomly distorted to produce a file of 20,000 unique stimuli. Each stimulus was converted into 16 primitive numerical attributes (statistical moments and edge counts) which were then scaled to fit into a range of integer values from 0 through 15. We typically train on the first 16000 items and then use the resulting model to predict the letter category for the remaining 4000. See the article cited above for more details.

5. Number of Instances: 20000

6. Number of Attributes: 17 (Letter category and 16 numeric features)

7. Attribute Information:

| | | | |
|---|---|---|---|
| 1. | lettr | capital letter | (26 values from A to Z) |
| 2. | x-box | horizontal position of box | (integer) |
| 3. | y-box | vertical position of box | (integer) |
| 4. | width | width of box | (integer) |
| 5. | high | height of box | (integer) |
| 6. | onpix | total # on pixels | (integer) |
| 7. | x-bar | mean x of on pixels in box | (integer) |
| 8. | y-bar | mean y of on pixels in box | (integer) |
| 9. | x2bar | mean x variance | (integer) |
| 10. | y2bar | mean y variance | (integer) |
| 11. | xybar | mean x y correlation | (integer) |
| 12. | x2ybr | mean of x * x * y | (integer) |
| 13. | xy2br | mean of x * y * y | (integer) |
| 14. | x-ege | mean edge count left to right | (integer) |
| 15. | xegvy | correlation of x-ege with y | (integer) |
| 16. | y-ege | mean edge count bottom to top | (integer) |
| 17. | yegvx | correlation of y-ege with x | (integer) |

8. Missing Attribute Values: None

9. Class Distribution:

| | | | | | | |
|---|---|---|---|---|---|---|
| 789 A | 766 B | 736 C | 805 D | 768 E | 775 F | 773 G |
| 734 H | 755 I | 747 J | 739 K | 761 L | 792 M | 783 N |
| 753 O | 803 P | 783 Q | 758 R | 748 S | 796 T | 813 U |
| 764 V | 752 W | 787 X | 786 Y | 734 Z | | |