|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **Letter Recognition Data Set**  *Download*: [Data Folder](http://archive.ics.uci.edu/ml/machine-learning-databases/letter-recognition/), [Data Set Description](http://archive.ics.uci.edu/ml/machine-learning-databases/letter-recognition/letter-recognition.names)  **Abstract**: Database of character image features; try to identify the letter |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 20000 | **Area:** | Computer | | **Attribute Characteristics:** | Integer | **Number of Attributes:** | 16 | **Date Donated** | 1991-01-01 | | **Associated Tasks:** | Classification | **Missing Values?** | No | **Number of Web Hits:** | 280890 |   **Source:**  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  **Data Set 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.  **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) |