!!! NB. DO NOT USE CROSS-VALIDATION WITH THIS DATASET !!!

Just train and test only once with the above

training and test sets.

PURPOSE

The database consists of the multi-spectral values

of pixels in 3x3 neighborhoods in a satellite image,

and the classification associated with the central pixel

in each neighborhood. The aim is to predict this

classification, given the multi-spectral values. In

the sample database, the class of a pixel is coded as

a number.

PROBLEM TYPE

Classification

AVAILABLE

This database was generated from Landsat Multi-Spectral

Scanner image data. These and other forms of remotely

sensed imagery can be purchased at a price from relevant

governmental authorities. The data is usually in binary

form, and distributed on magnetic tape(s).

SOURCE

The small sample database was provided by:

Ashwin Srinivasan

Department of Statistics and Modelling Science

University of Strathclyde

Glasgow

Scotland

UK

DESCRIPTION

One frame of Landsat MSS imagery consists of four digital images

of the same scene in different spectral bands. Two of these are

in the visible region (corresponding approximately to green and

red regions of the visible spectrum) and two are in the (near)

infra-red. Each pixel is a 8-bit binary word, with 0 corresponding

to black and 255 to white. The spatial resolution of a pixel is

about 80m x 80m.

The database is a (tiny) area of a scene, consisting of 82 x 100

pixels. Each line of data corresponds to a 3x3 square neighborhood

of pixels completely contained within the 82x100 sub-area. Each

line contains the pixel values in the four spectral bands

(converted to ASCII) of each of the 9 pixels in the 3x3

neighborhood and a number indicating the classification label of

the central pixel.

The number is a code for the following classes:

Number Class

1 red soil

2 cotton crop

3 grey soil

4 damp grey soil

5 soil with vegetation stubble

6 mixture class (all types present)

7 very damp grey soil

NB. There are no examples with class 6 in this dataset.

The data is given in random order and certain lines of data

have been removed so you cannot reconstruct the original image

from this dataset.

In each line of data the four spectral values for the top-left

pixel are given first, followed by the four spectral values for

the top-middle pixel and then those for the top-right pixel,

and so on with the pixels read out in sequence left-to-right and

top-to-bottom. Thus, the four spectral values for the central

pixel are given by attributes 17,18,19 and 20. If you like you

can use only these four attributes, while ignoring the others.

This avoids the problem which arises when a 3x3 neighborhood

straddles a boundary.

NUMBER OF EXAMPLES

training set 4435

test set 2000

NUMBER OF ATTRIBUTES

36 (= 4 spectral bands x 9 pixels in neighborhood)

ATTRIBUTES

The attributes are numerical, in the range 0 to 255.

CLASS

There are 6 decision classes: 1,2,3,4,5 and 7.

NB. There are no examples with class 6 in this dataset-

they have all been removed because of doubts about the

validity of this class.

AUTHOR

Ashwin Srinivasan

Department of Statistics and Data Modeling

University of Strathclyde

Glasgow

Scotland

UK