1. Title: Iris Plants Database

Updated Sept 21 by C.Blake - Added discrepency information

2. Sources:

(a) Creator: R.A. Fisher

(b) Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)

(c) Date: July, 1988

3. Past Usage:

- Publications: too many to mention!!! Here are a few.

1. Fisher,R.A. "The use of multiple measurements in taxonomic problems"

Annual Eugenics, 7, Part II, 179-188 (1936); also in "Contributions

to Mathematical Statistics" (John Wiley, NY, 1950).

2. Duda,R.O., & Hart,P.E. (1973) Pattern Classification and Scene Analysis.

(Q327.D83) John Wiley & Sons. ISBN 0-471-22361-1. See page 218.

3. Dasarathy, B.V. (1980) "Nosing Around the Neighborhood: A New System

Structure and Classification Rule for Recognition in Partially Exposed

Environments". IEEE Transactions on Pattern Analysis and Machine

Intelligence, Vol. PAMI-2, No. 1, 67-71.

-- Results:

-- very low misclassification rates (0% for the setosa class)

4. Gates, G.W. (1972) "The Reduced Nearest Neighbor Rule". IEEE

Transactions on Information Theory, May 1972, 431-433.

-- Results:

-- very low misclassification rates again

5. See also: 1988 MLC Proceedings, 54-64. Cheeseman et al's AUTOCLASS II

conceptual clustering system finds 3 classes in the data.

4. Relevant Information:

--- This is perhaps the best known database to be found in the pattern

recognition literature. Fisher's paper is a classic in the field

and is referenced frequently to this day. (See Duda & Hart, for

example.) The data set contains 3 classes of 50 instances each,

where each class refers to a type of iris plant. One class is

linearly separable from the other 2; the latter are NOT linearly

separable from each other.

--- Predicted attribute: class of iris plant.

--- This is an exceedingly simple domain.

--- This data differs from the data presented in Fishers article

(identified by Steve Chadwick, spchadwick@espeedaz.net )

The 35th sample should be: 4.9,3.1,1.5,0.2,"Iris-setosa"

where the error is in the fourth feature.

The 38th sample: 4.9,3.6,1.4,0.1,"Iris-setosa"

where the errors are in the second and third features.

5. Number of Instances: 150 (50 in each of three classes)

6. Number of Attributes: 4 numeric, predictive attributes and the class

7. Attribute Information:

1. sepal length in cm

2. sepal width in cm

3. petal length in cm

4. petal width in cm

5. class:

-- Iris Setosa

-- Iris Versicolour

-- Iris Virginica

8. Missing Attribute Values: None

Summary Statistics:

Min Max Mean SD Class Correlation

sepal length: 4.3 7.9 5.84 0.83 0.7826

sepal width: 2.0 4.4 3.05 0.43 -0.4194

petal length: 1.0 6.9 3.76 1.76 0.9490 (high!)

petal width: 0.1 2.5 1.20 0.76 0.9565 (high!)

9. Class Distribution: 33.3% for each of 3 classes.