'DESCR': '.. \_iris\_dataset:

Iris plants dataset

--------------------

\*\*Data Set Characteristics:\*\*

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

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

Attribute Information:

- sepal length in cm

- sepal width in cm

- petal length in cm

- petal width in cm

class:

- Iris-Setosa

- Iris-Versicolour

- Iris-Virginica

: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!)

============== ==== ==== ======= ===== ===================

:Missing Attribute Values: None

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

:Creator: R.A. Fisher

:Donor: Michael Marshall ([MARSHALL%PLU@io.arc.nasa.gov](mailto:MARSHALL%25PLU@io.arc.nasa.gov))

:Date: July, 1988

The famous Iris database, first used by Sir R.A. Fisher. The dataset is taken from Fisher\'s paper. Note that it\'s the same as in R, but not as in the UCI

Machine Learning Repository, which has two wrong data points.

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.

.. topic:: References

- 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).

- 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.

- 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.

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

on Information Theory, May 1972, 431-433.\n - See also: 1988 MLC Proceedings, 54-64. Cheeseman et al"s AUTOCLASS II\n conceptual clustering system finds 3 classes in the data.

- Many, many more ...',