

dataset

December 22, 2023

```
[1]: %pip install ucimlrepo -q
```

Note: you may need to restart the kernel to use updated packages.

```
[2]: from ucimlrepo import fetch_ucirepo
```

```
# fetch dataset
iris = fetch_ucirepo(id=53)

# data (as pandas dataframes)
X = iris.data.features
y = iris.data.targets

# metadata
print(iris.metadata)

# variable information
print(iris.variables)
```

```
{'uci_id': 53, 'name': 'Iris', 'repository_url':
'https://archive.ics.uci.edu/dataset/53/iris', 'data_url':
'https://archive.ics.uci.edu/static/public/53/data.csv', 'abstract': 'A small
classic dataset from Fisher, 1936. One of the earliest known datasets used for
evaluating classification methods.\n', 'area': 'Biology', 'tasks':
['Classification'], 'characteristics': ['Tabular'], 'num_instances': 150,
'num_features': 4, 'feature_types': ['Real'], 'demographics': [], 'target_col':
['class'], 'index_col': None, 'has_missing_values': 'no',
'missing_values_symbol': None, 'year_of_dataset_creation': 1936, 'last_updated':
'Tue Sep 12 2023', 'dataset_doi': '10.24432/C56C76', 'creators': ['R. A.
Fisher'], 'intro_paper': {'title': 'The Iris data set: In search of the source
of virginica', 'authors': 'A. Unwin, K. Kleinman', 'published_in':
'Significance, 2021', 'year': 2021, 'url': 'https://www.semanticscholar.org/pape
r/4599862ea877863669a6a8e63a3c707a787d5d7e', 'doi': '1740-9713.01589'},
'additional_info': {'summary': 'This is one of the earliest datasets used in the
literature on classification methods and widely used in statistics and machine
learning. 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.\n\nPredicted
attribute: class of iris plant.\n\nThis is an exceedingly simple domain.\n\nThis
```

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. ', 'purpose': 'N/A', 'funded_by': None, 'instances_represent': 'Each instance is a plant', 'recommended_data_splits': None, 'sensitive_data': None, 'preprocessing_description': None, 'variable_info': None, 'citation': None}}

	name	role	type	demographic \
0	sepal length	Feature	Continuous	None
1	sepal width	Feature	Continuous	None
2	petal length	Feature	Continuous	None
3	petal width	Feature	Continuous	None
4	class	Target	Categorical	None

	description	units	missing_values	
0		None	cm	no
1		None	cm	no
2		None	cm	no
3		None	cm	no
4	class of iris plant: Iris Setosa, Iris Versico...	None		no

```
[3]: print(X)
```

	sepal length	sepal width	petal length	petal width
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
..
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

[150 rows x 4 columns]

```
[4]: y
```

```
[4]:
      class
0    Iris-setosa
1    Iris-setosa
2    Iris-setosa
3    Iris-setosa
4    Iris-setosa
..      ...
```

```

145 Iris-virginica
146 Iris-virginica
147 Iris-virginica
148 Iris-virginica
149 Iris-virginica

[150 rows x 1 columns]

```

```
[5]: import pandas as pd
```

```
[6]: df = pd.concat(objs=[X, y], axis="columns")
df.head(n=10)
```

```
[6]:
```

	sepal length	sepal width	petal length	petal width	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa

```
[7]: import seaborn as sns
```

```
[8]: iris = sns.load_dataset("iris")
iris.head(n=10)
```

```
[8]:
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
5	5.4	3.9	1.7	0.4	setosa
6	4.6	3.4	1.4	0.3	setosa
7	5.0	3.4	1.5	0.2	setosa
8	4.4	2.9	1.4	0.2	setosa
9	4.9	3.1	1.5	0.1	setosa