



Iris データセット

URL: <https://www.kkaneko.jp/tools/bigdata/index.html>

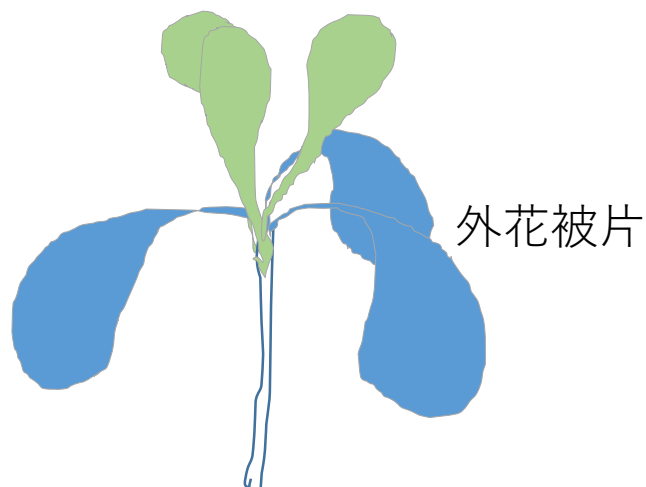
金子邦彦



アヤメ属 (Iris)



内花被片



外花被片

- 多年草
- 世界に 150種. 日本に 9種.
- 花被片は 6個
- **外花被片** (がいかひへん) **Sepal**
3個 (大型で下に垂れる)
- **内花被片** (ないかひへん) **Petal**
3個 (直立する)

Iris データセット



Iris データセット (データ数は 50×3)
のうち、先頭 10 行

sepal_length	sepal_width	petal_length	petal_width	species
5.1	3.5	1.4	0.2	setosa
4.9	3	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa

外花被片 (Sepal) の長さ と **幅**
内花被片 (Petal) の長さ と **幅**
種類

特徴量

ラベル

◆ **3種のアヤメの外花被**
辺、内花被片を計測

◆ 種類も記録

setosa

versicolor

virginica

◆ データ数は **50×3**

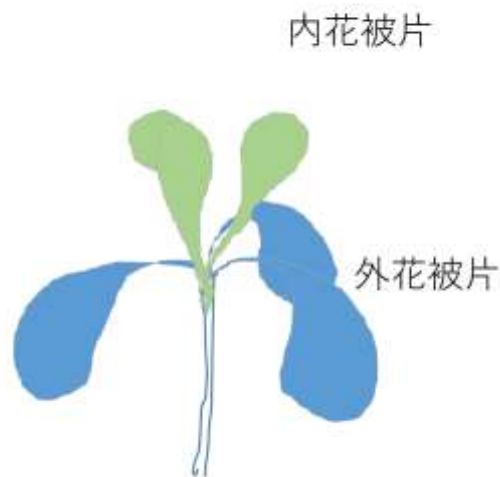
作成者 : Ronald Fisher

作成年 : 1936

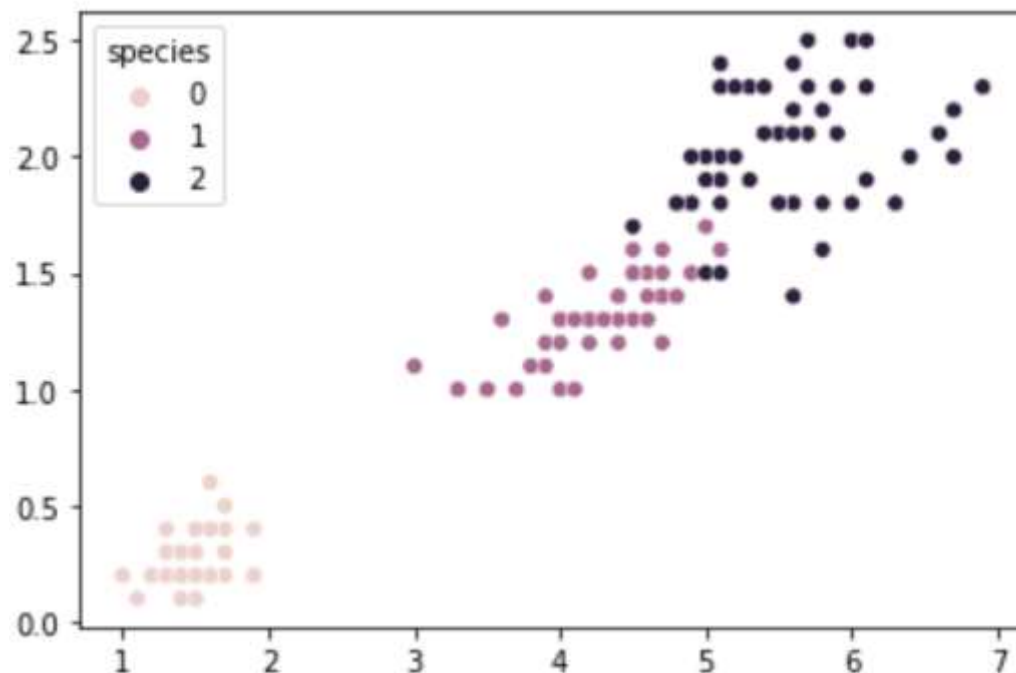
Iris データセットの散布図



アヤメ属 (Iris)



縦軸：内花被片の幅



横軸：内花被片の長さ

次の **3種類** の分類済みのデータ

0: **setosa**

1: **versicolor**

2: **virginica**

Iris データセットと配列 (アレイ)



Iris データセット

sepal_length	sepal_width	petal_length	petal_width	species
5.1	3.5	1.4	0.2	setosa
4.9	3	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa



```
[[5.1 3.5 1.4 0.2]
 [4.9 3.  1.4 0.2]
 [4.7 3.2 1.3 0.2]
 [4.6 3.1 1.5 0.2]
 [5.  3.6 1.4 0.2]
 [5.4 3.9 1.7 0.4]
 [4.6 3.4 1.4 0.3]
 [5.  3.4 1.5 0.2]
 [4.4 2.9 1.4 0.2]
 [4.9 3.1 1.5 0.1]
 [5.4 3.7 1.5 0.2]]
```

```
[0
 0
 0
 0
 0
 0
 0
 0
 0
 0
 0]
```

特徴量 (数値)
サイズ : **150 × 4**

ラベル (数値)
サイズ : **150**

```
setosa → 0
versicolor → 1
virginia → 2
```

ラベルの数値化

Iris データセットは Python でも利用可能



ソースコード

```
from sklearn.datasets import load_iris
iris = load_iris()
X, y = iris.data, iris.target
print(X)
print(y)
```

```
コマンド プロンプト - python
Microsoft Windows [Version 10.0.25197.1000]
(c) Microsoft Corporation. All rights reserved.

C:\Users\me>python
Python 3.10.7 (tags/v3.10.7:6cc6b13, Sep 5 2022, 14:08:36) [MSC v.1933 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>> from sklearn.datasets import load_iris
>>> iris = load_iris()
>>> X, y = iris.data, iris.target
>>> print(X)
[[5.1 3.5 1.4 0.2]
 [4.9 3. 1.4 0.2]
 [4.7 3.2 1.3 0.2]
 [4.6 3.1 1.5 0.2]
 [5. 3.6 1.4 0.2]
 [5.4 3.9 1.7 0.4]
 [4.6 3.4 1.4 0.3]
 [5. 3.4 1.5 0.2]
 [4.4 2.9 1.4 0.2]
```