



## 6.5 鸢尾花数据集可视化



## 6.5.1 下载鸢尾花数据集

### ■ 鸢尾花



### ■ Iris 数据集

- Anderson's Iris Data Set
  - 1936, Ronald Fisher, *The use of multiple measurements in taxonomic problems*
- |              |                 |
|--------------|-----------------|
| ■ 加拿大的加斯帕半岛  | ■ 3种鸢尾花类别       |
| ■ 同一天的同一个时间段 | ■ 每个类别有50个样本    |
| ■ 在相同的牧场上    | ■ 每个样本中包括4种鸢尾花的 |
| ■ 由同一个人      | 属性特征, 和鸢尾花的品种   |
| ■ 使用相同的测量仪器  |                 |



### Iris数据集

- 150个样本
- 4个属性
  - 花萼长度 (Sepal Length)
  - 花萼宽度 (Sepal Width)
  - 花瓣长度 (Petal Length)
  - 花瓣宽度 (Petal Width)
- 1个标签
  - 山鸢尾 (Setosa)
  - 变色鸢尾 (Versicolour)
  - 维吉尼亚鸢尾 (Virginica)

Iris数据集示例

| Sepal length | Sepal width | Petal length | Petal width | Species       |
|--------------|-------------|--------------|-------------|---------------|
| 5.1          | 3.5         | 1.4          | 0.2         | I. setosa     |
| 4.9          | 3           | 1.4          | 0.2         | I. setosa     |
| 4.7          | 3.2         | 1.3          | 0.2         | I. setosa     |
| 4.6          | 3.1         | 1.5          | 0.2         | I. setosa     |
| 5            | 3.6         | 1.4          | 0.3         | I. setosa     |
| 7            | 3.2         | 4.7          | 1.4         | I. versicolor |
| 6.4          | 3.2         | 4.5          | 1.5         | I. versicolor |
| 6.9          | 3.1         | 4.9          | 1.5         | I. versicolor |
| 6.3          | 3.3         | 6            | 2.5         | I. virginica  |
| 5.8          | 2.7         | 5.1          | 1.9         | I. virginica  |
| 7.1          | 3           | 5.9          | 2.1         | I. virginica  |
| 6.3          | 2.9         | 5.6          | 1.8         | I. virginica  |
| 6.5          | 3           | 5.8          | 2.2         | I. virginica  |



### □ get\_file()函数——下载数据集

```
tf.keras.utils.get_file(fname, origin, cache_dir)
```

#### ■ 参数

fname: 下载后的文件名;

origin: 文件的URL地址;

cache\_dir: 下载后文件的存储位置。C:\Users\当前用户名\.keras\datasets

#### ■ 返回值: 下载后的文件在本地磁盘中的绝对路径。



### □ 下载鸢尾花数据集 iris

**训练数据集**          iris\_training.csv    120条数据

**测试数据集**          iris\_test.csv        30条数据

```
TRAIN_URL = "http://download.tensorflow.org/data/iris_training.csv"  
train_path = tf.keras.utils.get_file("iris_training.csv", TRAIN_URL)
```

```
Downloading data from http://download.tensorflow.org/data/iris_training.csv  
8192/2194 [=====] - 1s 66us/step
```

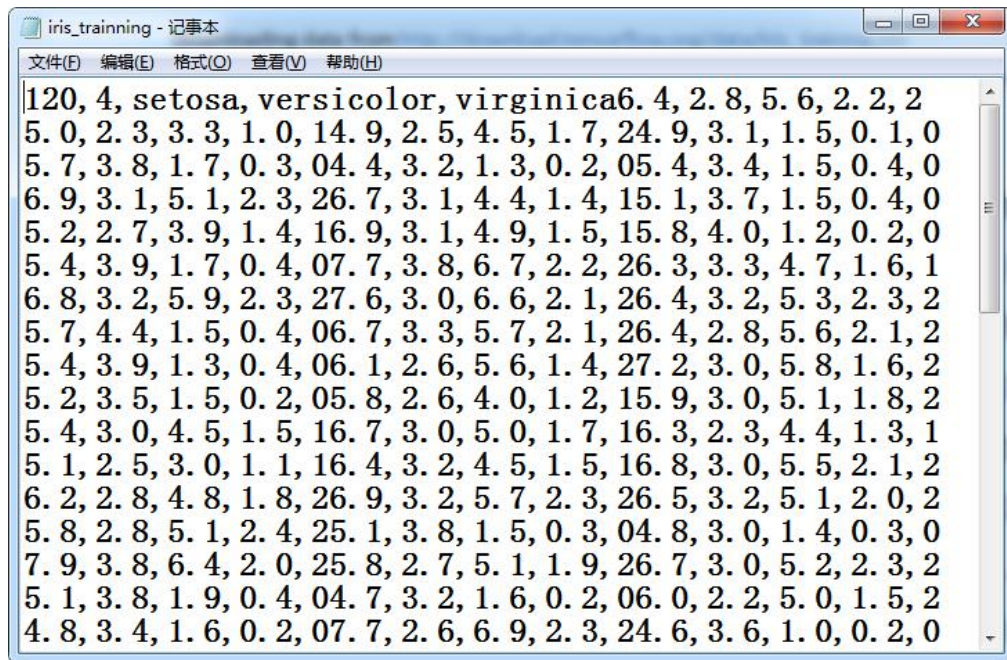
C:\Users\Administrator\.keras\datasets\iris\_training.csv





## 6.5.1 下载鸢尾花数据集

### □ csv文件



```
iris_training - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)
120, 4, setosa, versicolor, virginica
6.4, 2.8, 5.6, 2.2, 2
5.0, 2.3, 3.3, 1.0, 14.9, 2.5, 4.5, 1.7, 24.9, 3.1, 1.5, 0.1, 0
5.7, 3.8, 1.7, 0.3, 04.4, 3.2, 1.3, 0.2, 05.4, 3.4, 1.5, 0.4, 0
6.9, 3.1, 5.1, 2.3, 26.7, 3.1, 4.4, 1.4, 15.1, 3.7, 1.5, 0.4, 0
5.2, 2.7, 3.9, 1.4, 16.9, 3.1, 4.9, 1.5, 15.8, 4.0, 1.2, 0.2, 0
5.4, 3.9, 1.7, 0.4, 07.7, 3.8, 6.7, 2.2, 26.3, 3.3, 4.7, 1.6, 1
6.8, 3.2, 5.9, 2.3, 27.6, 3.0, 6.6, 2.1, 26.4, 3.2, 5.3, 2.3, 2
5.7, 4.4, 1.5, 0.4, 06.7, 3.3, 5.7, 2.1, 26.4, 2.8, 5.6, 2.1, 2
5.4, 3.9, 1.3, 0.4, 06.1, 2.6, 5.6, 1.4, 27.2, 3.0, 5.8, 1.6, 2
5.2, 3.5, 1.5, 0.2, 05.8, 2.6, 4.0, 1.2, 15.9, 3.0, 5.1, 1.8, 2
5.4, 3.0, 4.5, 1.5, 16.7, 3.0, 5.0, 1.7, 16.3, 2.3, 4.4, 1.3, 1
5.1, 2.5, 3.0, 1.1, 16.4, 3.2, 4.5, 1.5, 16.8, 3.0, 5.5, 2.1, 2
6.2, 2.8, 4.8, 1.8, 26.9, 3.2, 5.7, 2.3, 26.5, 3.2, 5.1, 2.0, 2
5.8, 2.8, 5.1, 2.4, 25.1, 3.8, 1.5, 0.3, 04.8, 3.0, 1.4, 0.3, 0
7.9, 3.8, 6.4, 2.0, 25.8, 2.7, 5.1, 1.9, 26.7, 3.0, 5.2, 2.3, 2
5.1, 3.8, 1.9, 0.4, 04.7, 3.2, 1.6, 0.2, 06.0, 2.2, 5.0, 1.5, 2
4.8, 3.4, 1.6, 0.2, 07.7, 2.6, 6.9, 2.3, 24.6, 3.6, 1.0, 0.2, 0
```





## 6.5.1 下载鸢尾花数据集

### ■ iris\_training.csv

- 120个样本
- 4个属性
  - 花萼长度 (Sepal Length)
  - 花萼宽度 (Sepal Width)
  - 花瓣长度 (Petal Length)
  - 花瓣宽度 (Petal Width)
- 1个标签
  - 0- 山鸢尾 (Setosa)
  - 1- 变色鸢尾 (Versicolour)
  - 2- 维吉尼亚鸢尾 (Virginica)

|    | A   | B   | C      | D          | E         |
|----|-----|-----|--------|------------|-----------|
| 1  | 120 | 4   | setosa | versicolor | virginica |
| 2  | 6.4 | 2.8 | 5.6    | 2.2        | 2         |
| 3  | 5   | 2.3 | 3.3    | 1          | 1         |
| 4  | 4.9 | 2.5 | 4.5    | 1.7        | 2         |
| 5  | 4.9 | 3.1 | 1.5    | 0.1        | 0         |
| 6  | 5.7 | 3.8 | 1.7    | 0.3        | 0         |
| 7  | 4.4 | 3.2 | 1.3    | 0.2        | 0         |
| 8  | 5.4 | 3.4 | 1.5    | 0.4        | 0         |
| 9  | 6.9 | 3.1 | 5.1    | 2.3        | 2         |
| 10 | 6.7 | 3.1 | 4.4    | 1.4        | 1         |
| 11 | 5.1 | 3.7 | 1.5    | 0.4        | 0         |
| 12 | 5.2 | 2.7 | 3.9    | 1.4        | 1         |
| 13 | 6.9 | 3.1 | 4.9    | 1.5        | 1         |
| 14 | 5.8 | 4   | 1.2    | 0.2        | 0         |
| 15 | 5.4 | 3.9 | 1.7    | 0.4        | 0         |



### ■ 下载数据集

```
TRAIN_URL = "http://download.tensorflow.org/data/iris_training.csv"  
train_path = tf.keras.utils.get_file("iris_training.csv", TRAIN_URL)
```

- split()函数: 通过**指定的分隔符**对字符串进行**切片**, 并返回一个**列表**。

```
TRAIN_URL = "http://download.tensorflow.org/data/iris_training.csv"  
TRAIN_URL.split('/')
```

运行结果:

```
['http:', '', 'download.tensorflow.org', 'data', 'iris_training.csv']
```



### □ 获取文件名

```
TRAIN_URL = "http://download.tensorflow.org/data/iris_training.csv"  
fname_list=TRAIN_URL.split('/')  
fname_list[-1]
```

```
TRAIN_URL.split('/')[-1]
```

运行结果：

```
'iris_training.csv '
```

### □ 下载数据集

下载其他数据集时，只要改变URL地址即可

```
TRAIN_URL = "http://download.tensorflow.org/data/iris_training.csv"  
train_path = tf.keras.utils.get_file(TRAIN_URL.split('/')[-1], TRAIN_URL)
```

自动获取数据集文件名

