# 1.安装scikit-learn

## 这里安装2024年的稳定版: pip3 install Scikit-Learn==1.5.2

|  |
| --- |
|  |

## 安装完毕用一个程序检查一下

|  |
| --- |
|  |

# 2.sklearn数据集

|  |
| --- |
|  |
|  |
|  |
|  |
|  |

## 获取到的数据集不能够全部用来训练模型,需要留一部分作为测试用

|  |
| --- |
|  |
|  |
| 注意返回值的顺序,分别对应下面的变量 |

在Windows中，sklearn下载数据集时没有指定数据保存地址，则该数据集会默认下载到你本机安装python的内置文件夹下的Lib/site-packages/sklearn/datasets/目录下,我们需要学习的数据集保存在Lib/site-packages/sklearn/datasets/data文件夹下面

|  |
| --- |
|  |

## 实例1

|  |
| --- |
| import sklearn from sklearn.datasets import load\_iris  # print(sklearn.\_\_version\_\_) iris\_data = load\_iris() # print("返回数据:",iris\_data) # print("返回数据的特征值:",iris\_data["data"]) # print("返回数据的目标值:",iris\_data.target) # print("返回数据的特性的名字:",iris\_data.feature\_names) # ['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal width (cm)'] # print("返回数据的目标值的名字:",iris\_data.target\_names) # 返回数据的目标值的名字: ['setosa' 'versicolor' 'virginica'] print("返回数据的描述:",iris\_data.DESCR) #很多内容 |

## 实例2.

|  |
| --- |
| import sklearn from sklearn.datasets import load\_iris from sklearn.model\_selection import train\_test\_split """  把数据集分割为训练集和测试集,调用train\_test\_split返回值的顺序是训练集的特征值,测试集的特征值,训练集的目标值,测试集的目标值分别用:  x\_train,x\_test,y\_train,y\_test来表示 """ # print(sklearn.\_\_version\_\_) iris\_data = load\_iris() x\_train,x\_test,y\_train,y\_test = train\_test\_split(iris\_data.data,iris\_data.target,  test\_size=0.25,random\_state=22) print("训练集的行数和列数",x\_train.shape) print("测试集的行数和列数",x\_test.shape) |