# 机器学习入门

* 概述

**AlphaGO vs 柯洁**

**Tesla(特斯拉)**





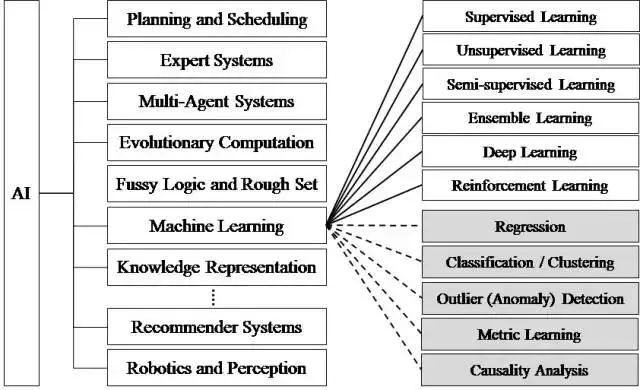
**人工智能实现方法**



**人工智能是一个梦想，希望通过计算机构造出复杂的、拥有与人类智慧同样本质特性的机器。**

**机器学习是实现人工智能的一种方法。**

**深度学习是实现机器学习的一种方法。**



**机器学习知识体系**

[**https://img-blog.csdn.net/20170729203658751?watermark/2/text/aHR0cDovL2Jsb2cuY3Nkbi5uZXQvZHJlYW1fYW4=/font/5a6L5L2T/fontsize/400/fill/I0JBQkFCMA==/dissolve/70/gravity/SouthEast**](https://img-blog.csdn.net/20170729203658751?watermark/2/text/aHR0cDovL2Jsb2cuY3Nkbi5uZXQvZHJlYW1fYW4=/font/5a6L5L2T/fontsize/400/fill/I0JBQkFCMA==/dissolve/70/gravity/SouthEast)

**机器学习项目流程**

**1.目标定义**

**a.任务理解**

**b.指标确定**

**c.实际问题抽象为数学问题**

**2.获取数据**

**a.爬虫或第三方数据源**

**b.数据质量把控(数据完整无缺，各项指标齐全)**

**3.数据清洗**

**a.重复数据**

**b.缺失值**

**c.异常值**

**d.数据类型转换**

**4.数据变换**

**a.属性变换**

**b.数据集成**

**c.归一化**

**5.探索性数据分析**

**a.变量识别**

**b.单变量分析**

**c.双变量分析**

**d.相关系数**

**6.模型构建**

**a.回归**

**b.分类**

**c.聚类**

**d.推荐**

**7.模型评测**

**a.多模型对比**

**b.模型优化**

**8.模型部署**

**常用分析工具**

**1.学术型**

**a.excel**

**b.matlab**

**c.R**

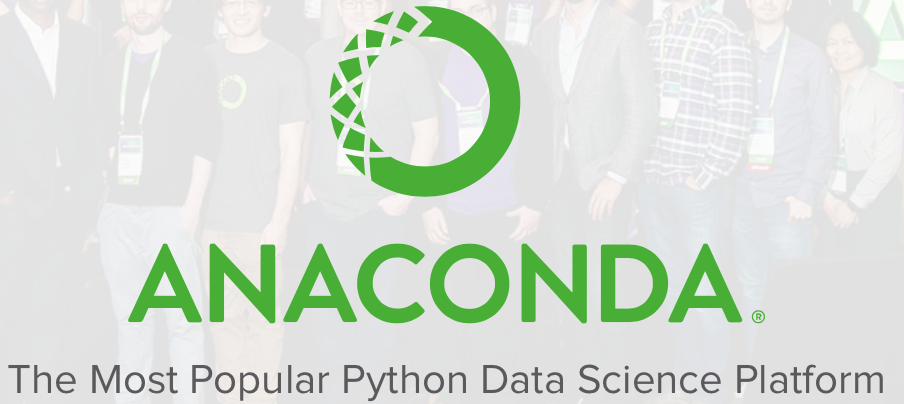
**d.SPSS**

**2.工程化**

**Python**

**Spark mllib**

**工程化工具推荐**

** **

* 预备环境

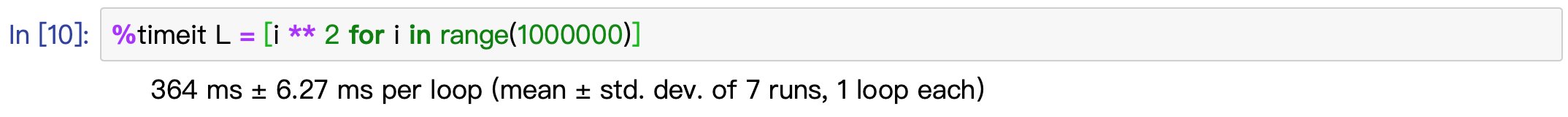
1. Python3 解释器
2. Anaconda安装文件

* Jupyter NoteBook基础

1. 基本使用
2. 魔法指令

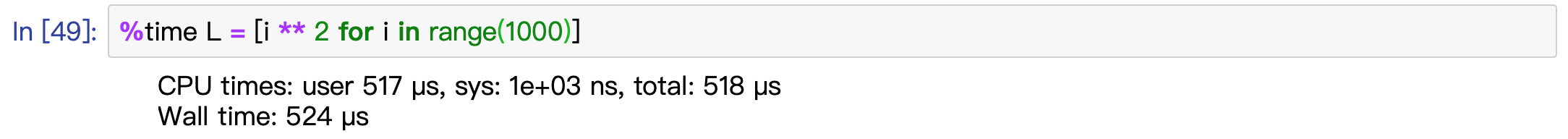
%run xxx.py 加载单一脚本

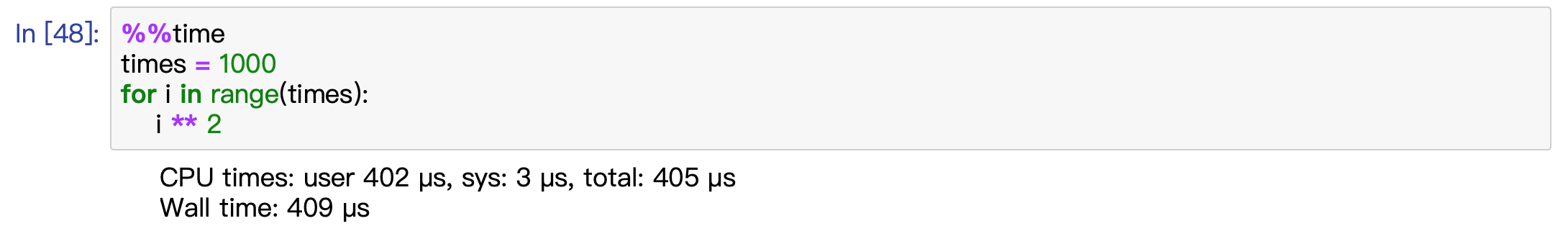
%timeit执行适当次数的程序代码，然后给出运行最快3次的平均值。



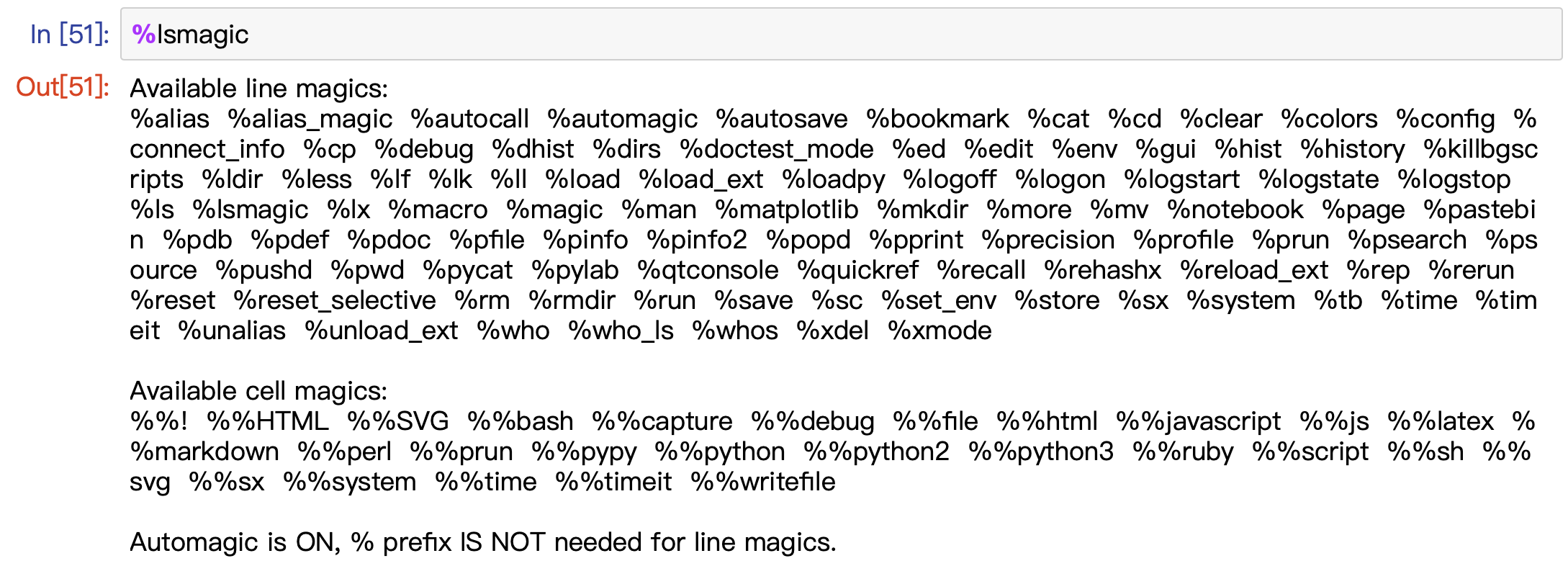


%time 代码执行一次的运行时间

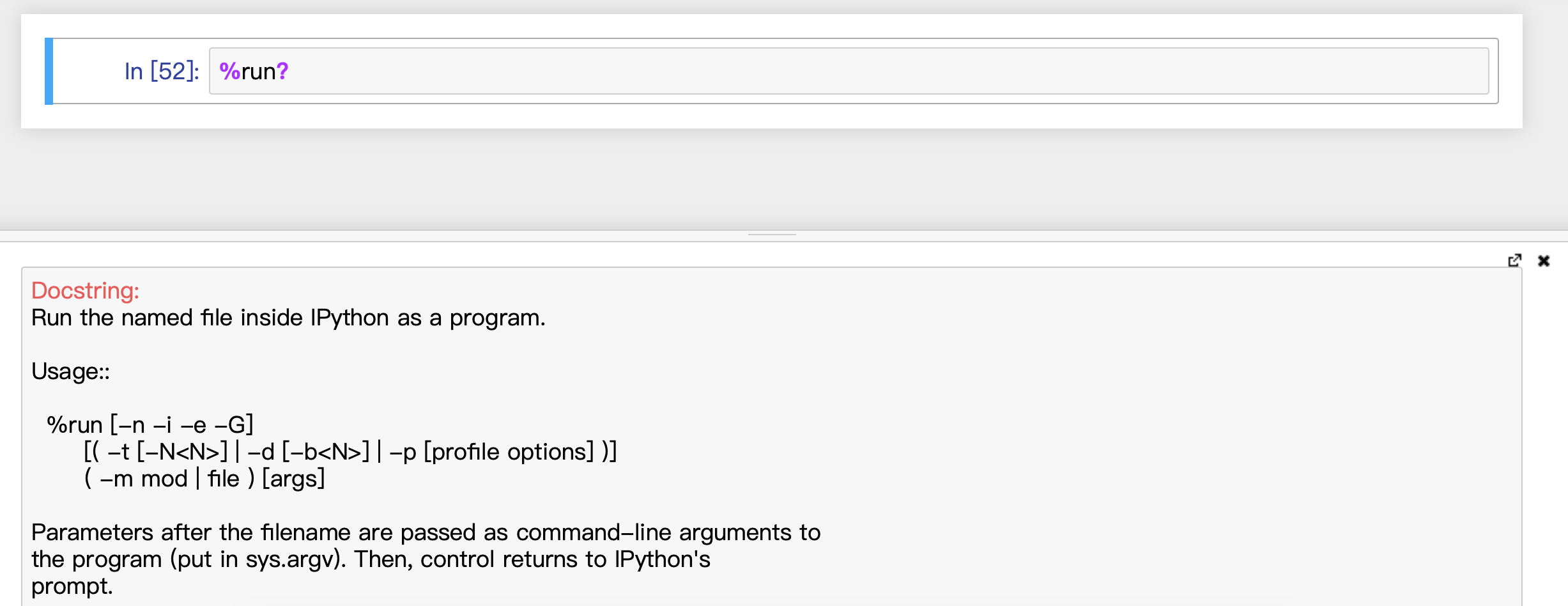




查看内置魔法命令



查看文档

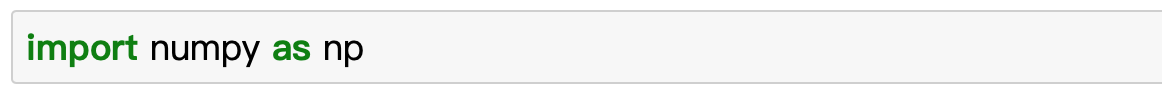


* Numpy基础

Numpy的核心就是Array，也就是数组。

通过创建多维数组，将数组看做矩阵来操作。

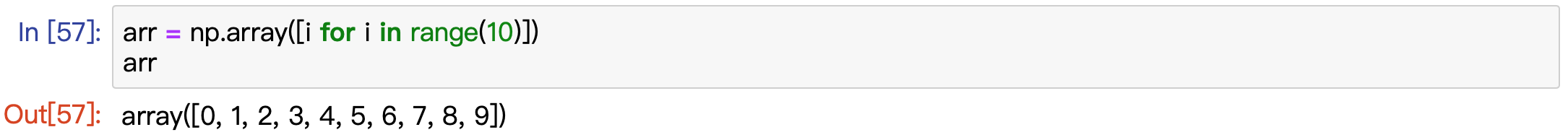
**导入Numpy**



**查看Numpy版本**



**创建Array**

****

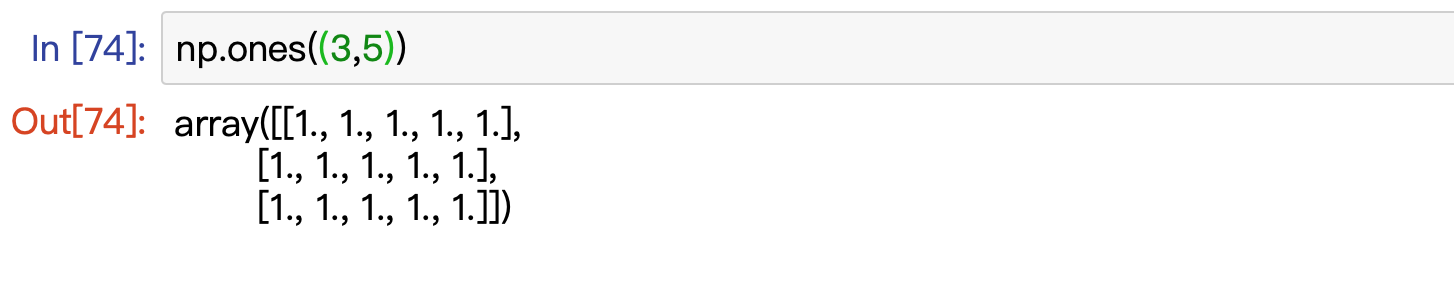
**查看元素类型**

****

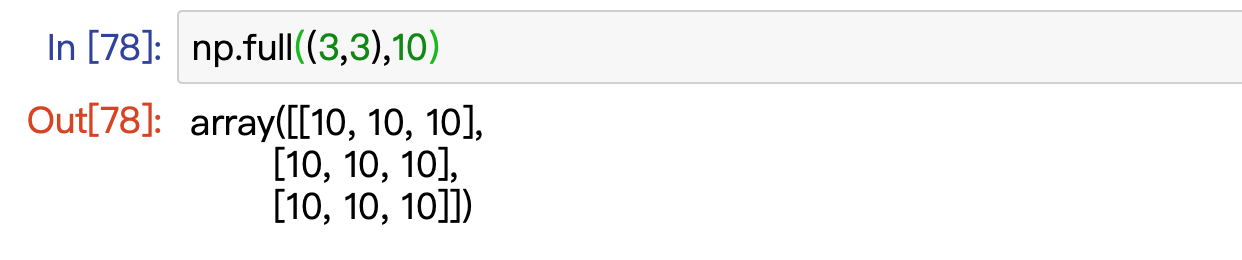
**创建零数组**

****

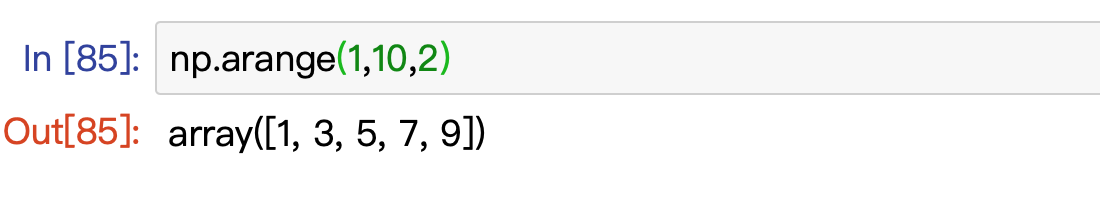
**创建全一数组**

****

**创建指定值矩阵**

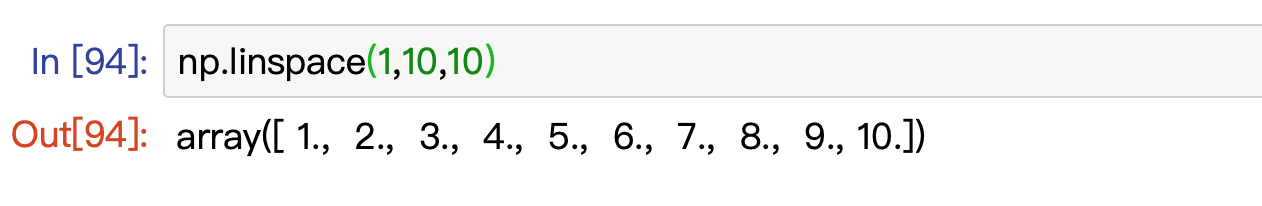
****

**arange**

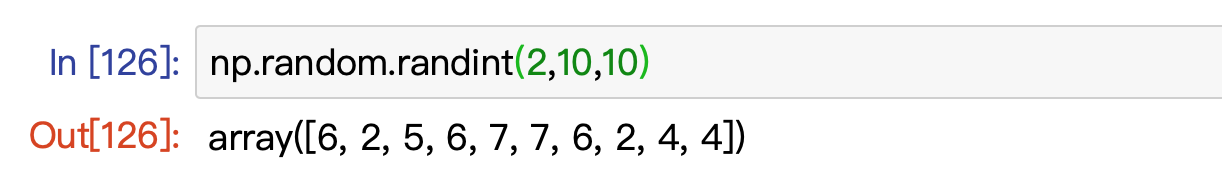
****

**与range()不同的是，arange()的步长可以使用浮点数**

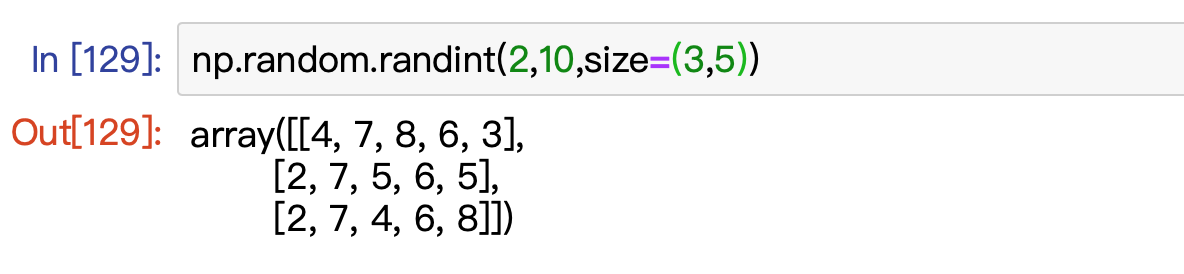
**linspace**

****

**random.randint**

****

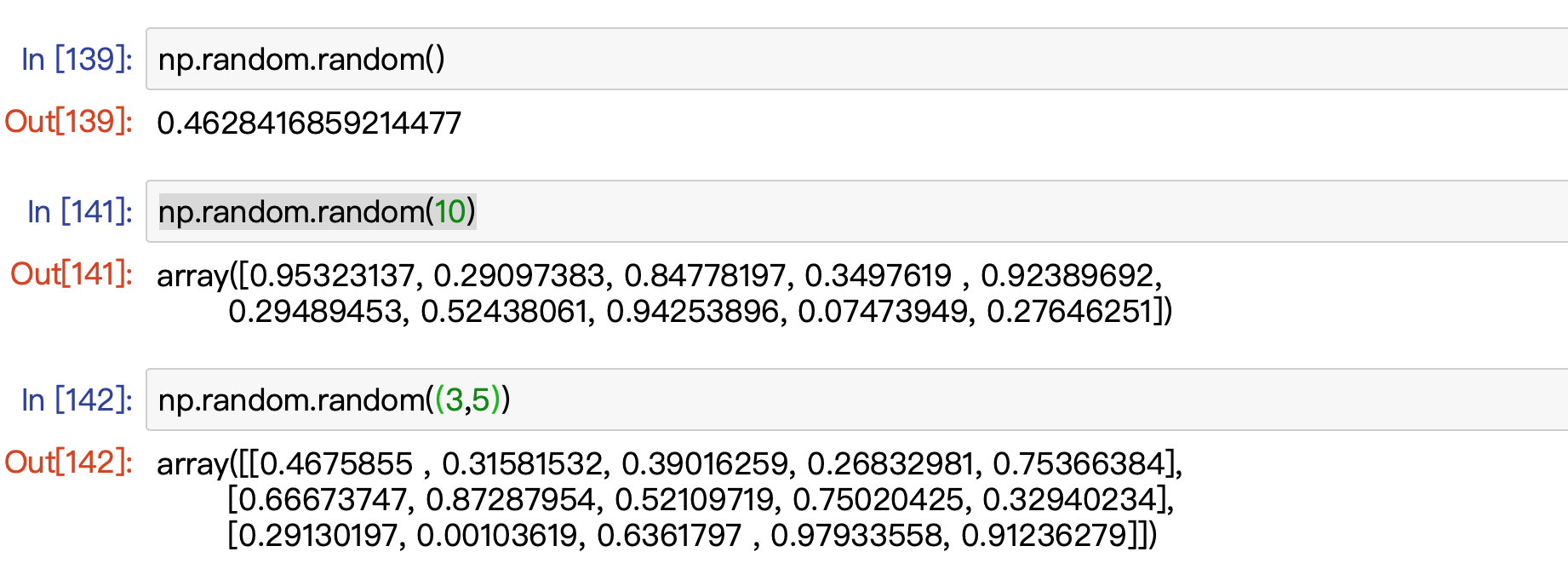
**第三个参数，表示随机生成的个数，返回一个数组**

****

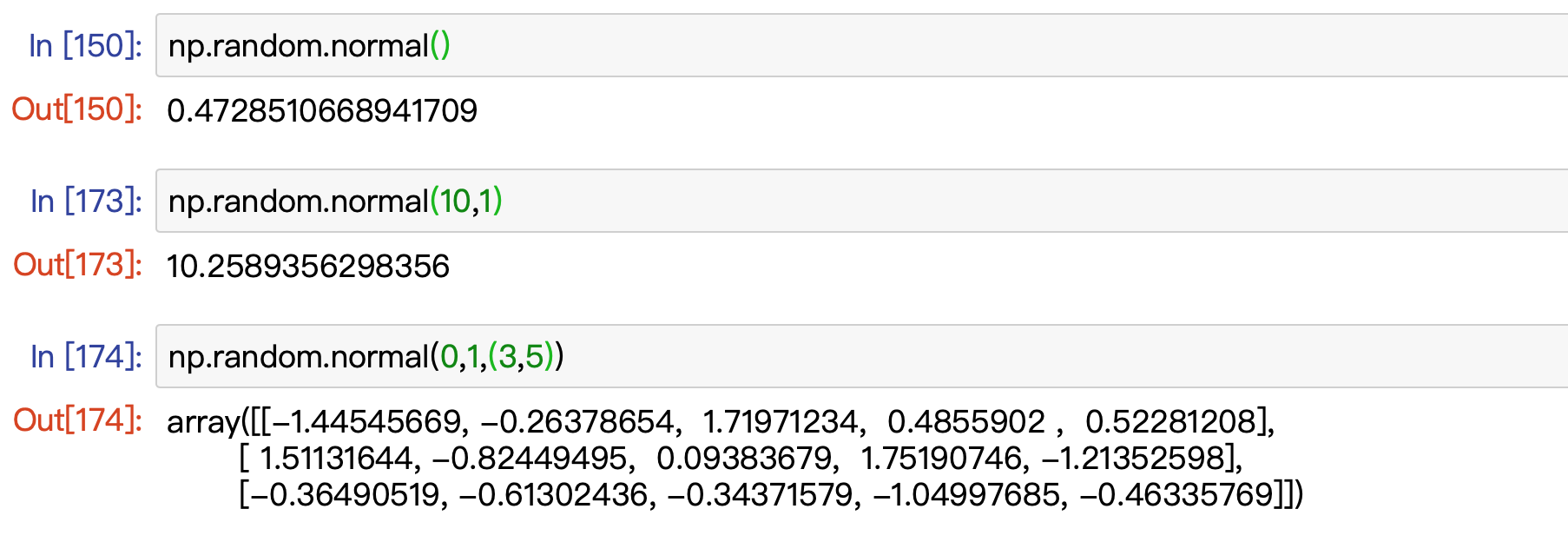
**random.seed(随机种子)**

****

**random.random()**

****

**random.normal(生成均值为0，方差为1符合正态分布的随机浮点数)**

****

**第一个参数为均值，第二个参数为方差**

* Numpy Array基本属性

**ndim 查看数组维数**



**shape 查看数组结构**



**size 查看数组元素个数**

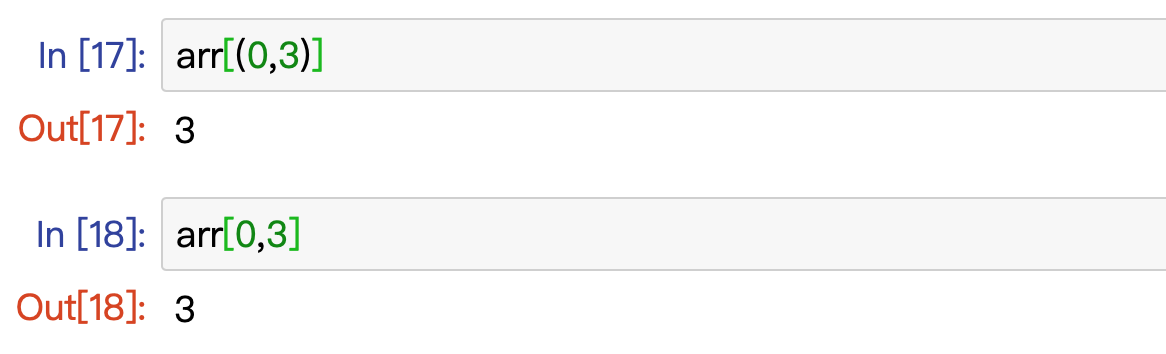


* Numpy Array数据访问

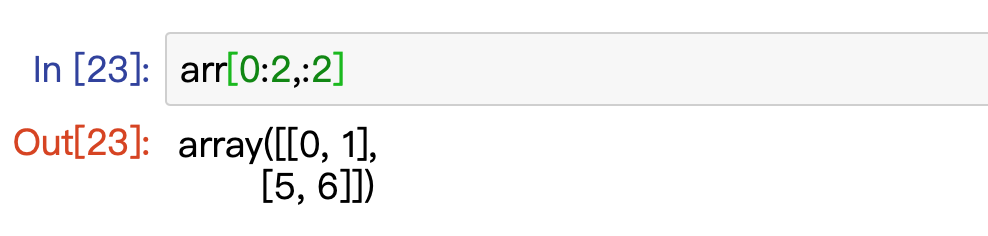
**根据下标访问**



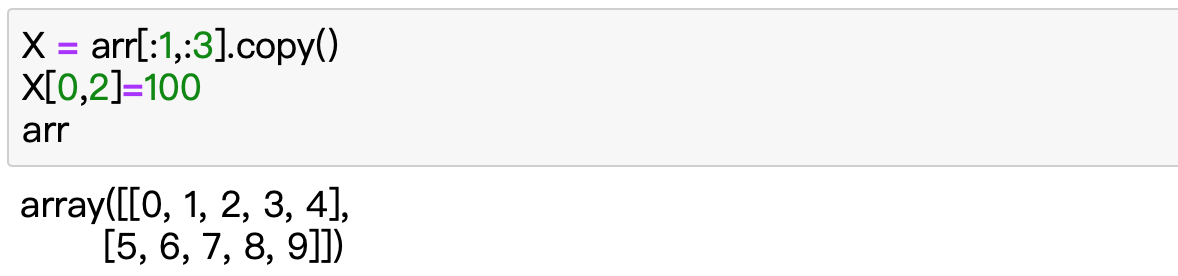
**元组访问**

****

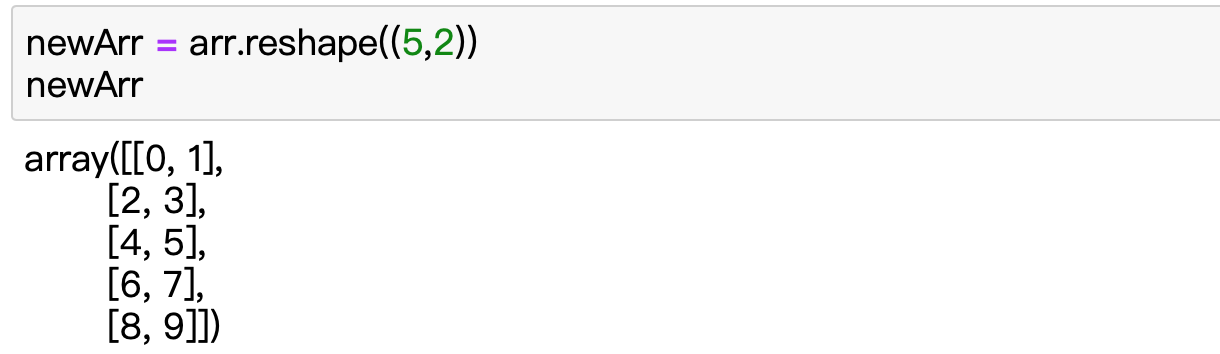
**切片访问**

****

**数组拷贝**

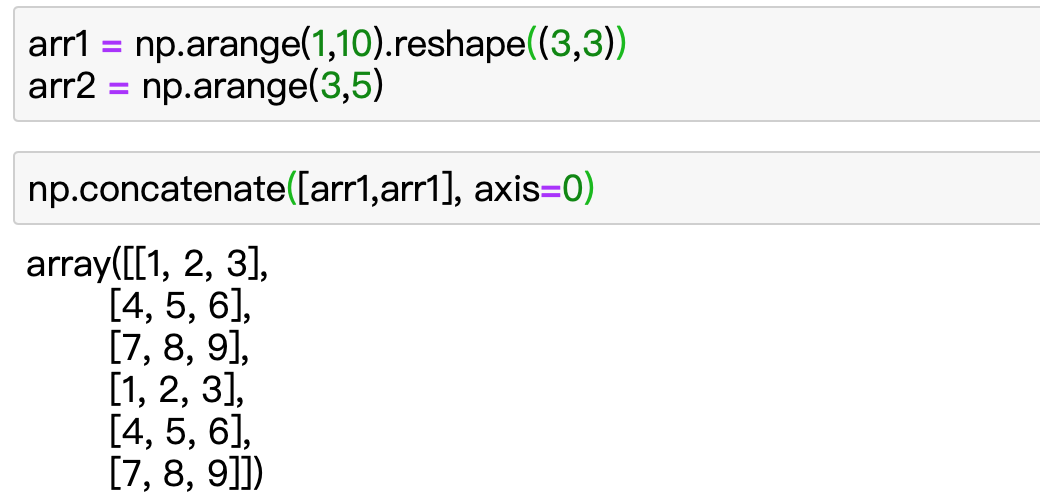
****

**数组维度改变**

****

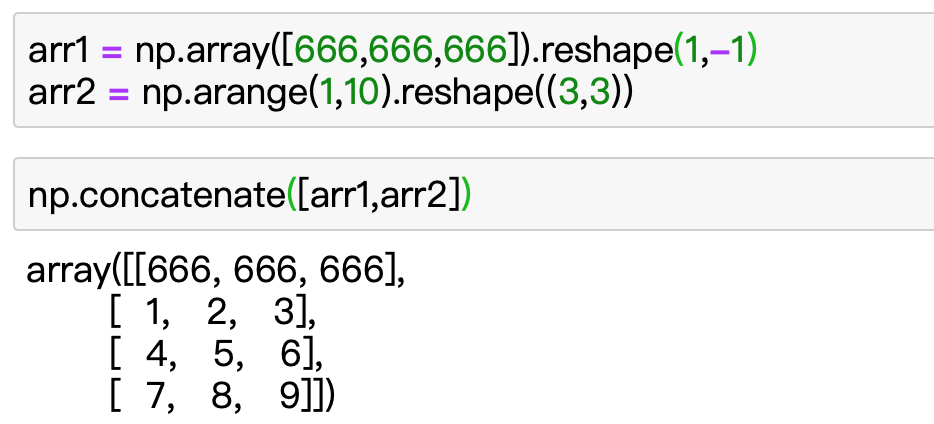
* Numpy Array合并与分割

**数组合并**

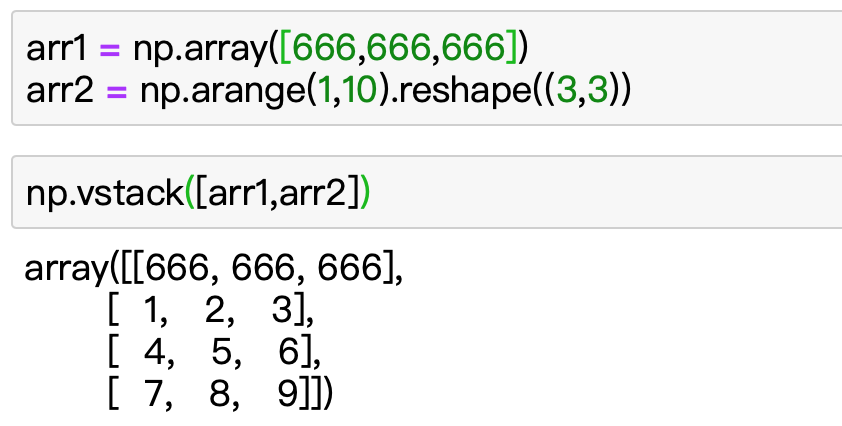
****

**axis=0表示按行合并，axis=1表示按列合并**

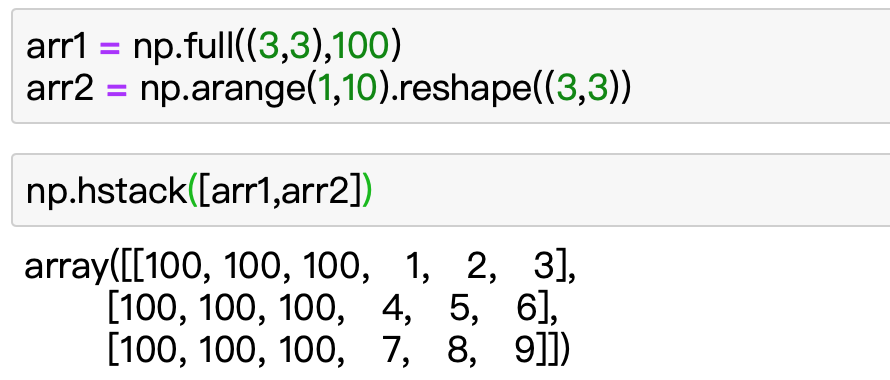
**不同维度的数组合并**

****

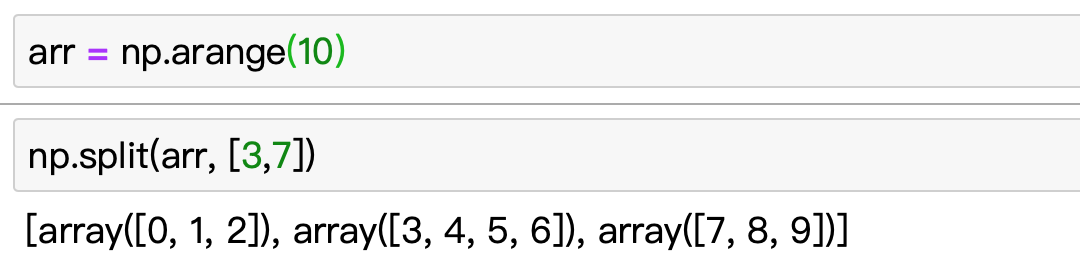
**垂直堆叠**

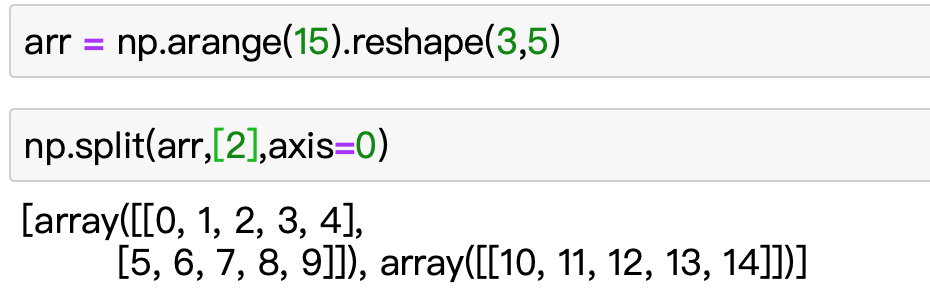
****

**水平堆叠**

****

**分割操作**

****

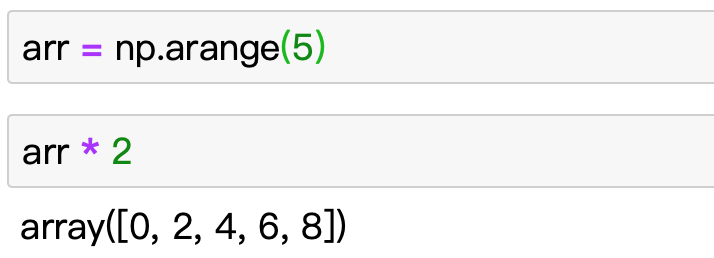
****

**水平与垂直分割**

****

* Numpy Array运算

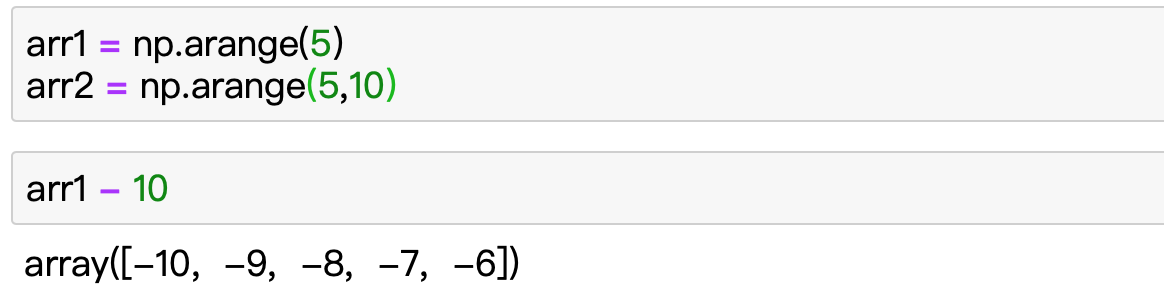
**向量数乘**

****

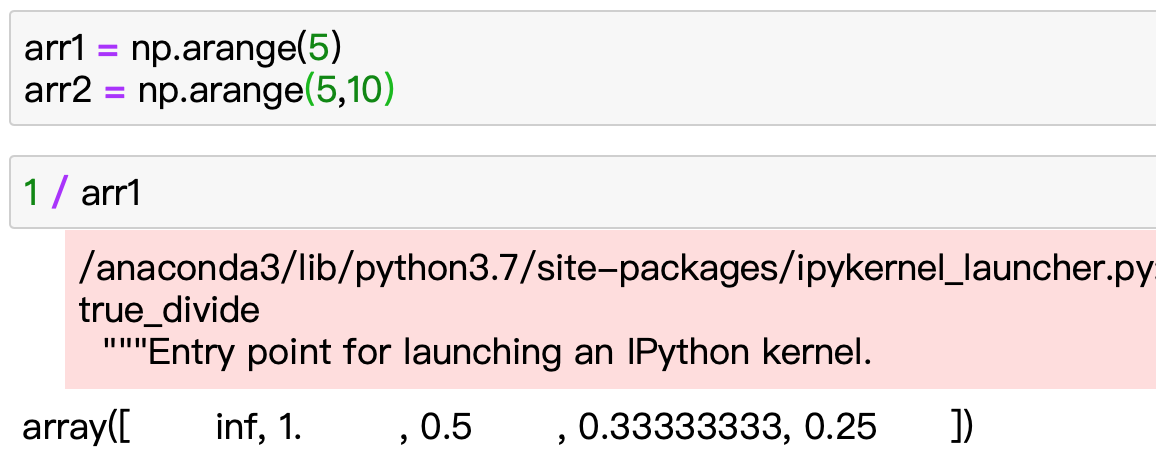
**向量加法**

****

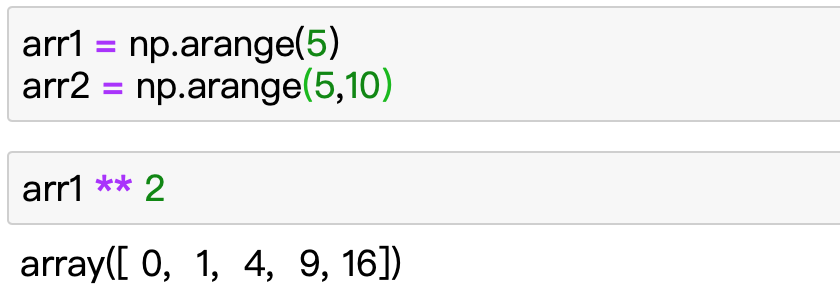
**向量减法**

****

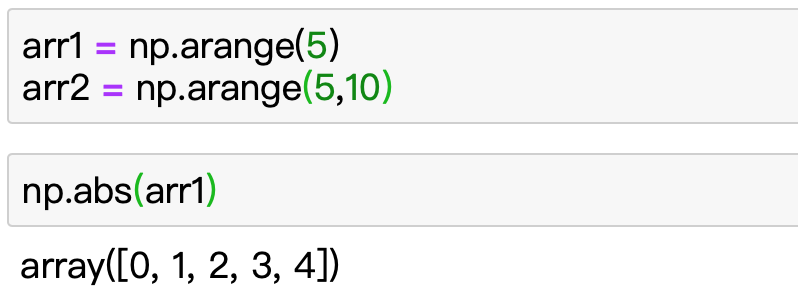
**向量倒数**

****

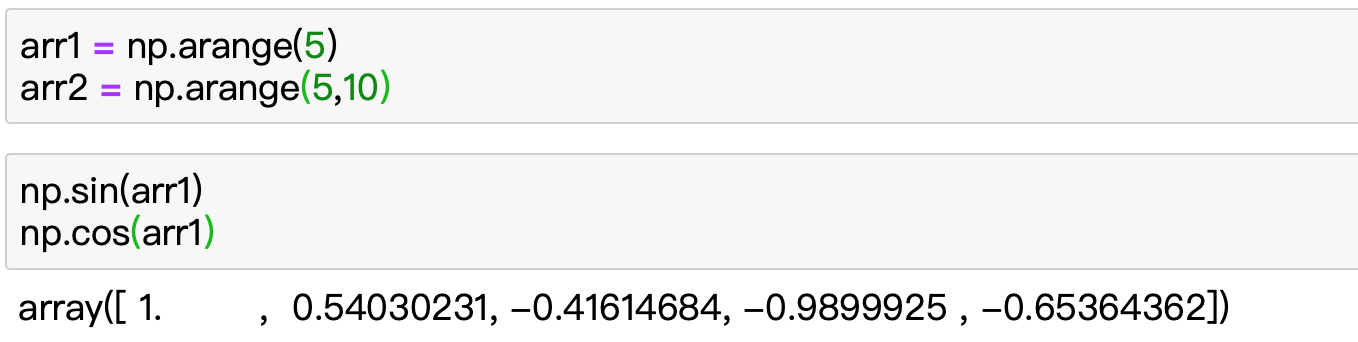
**向量次方**

****

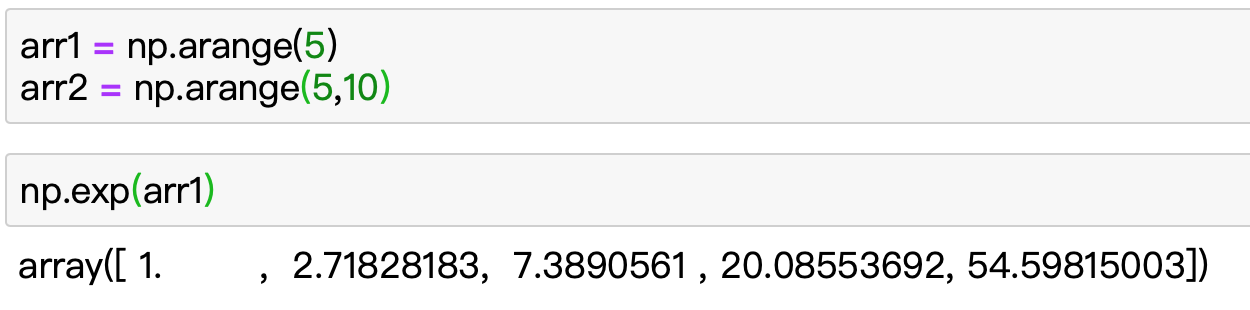
**向量绝对值**

****

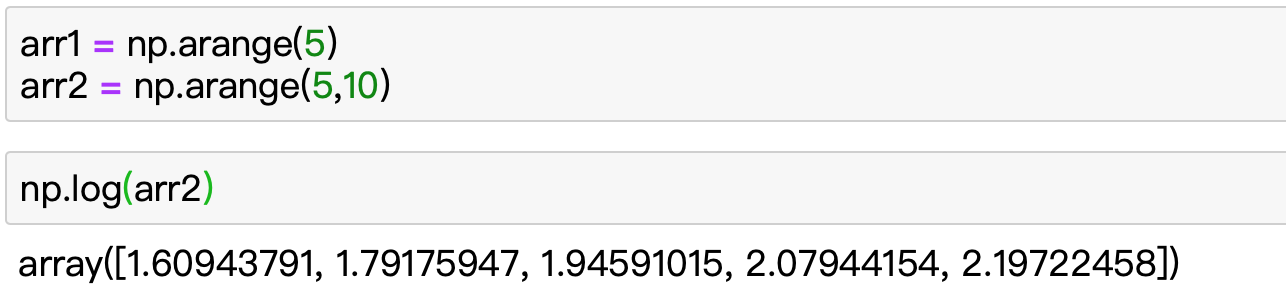
**向量三角函数**

****

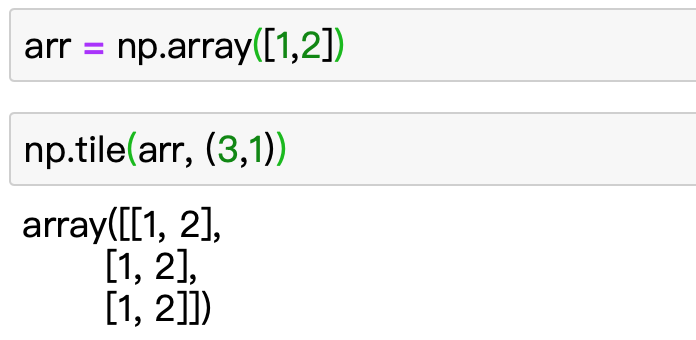
**向量exp**

****

**向量log**

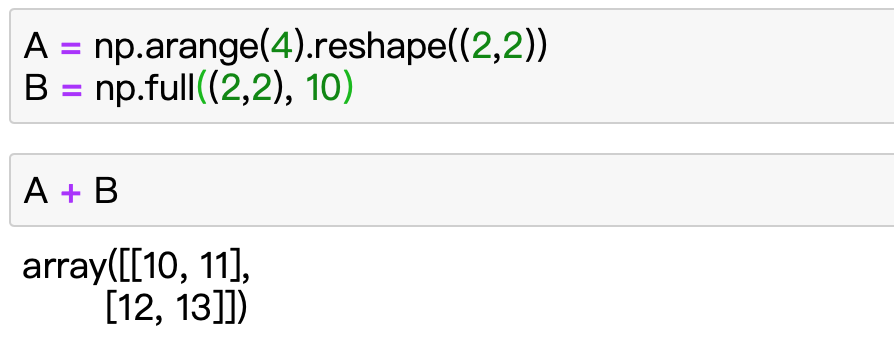
****

**向量堆叠**

****

* Numpy 矩阵运算

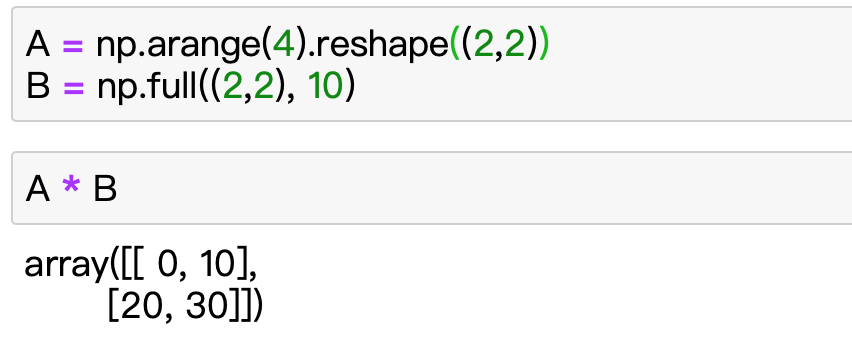
**矩阵求和**

****

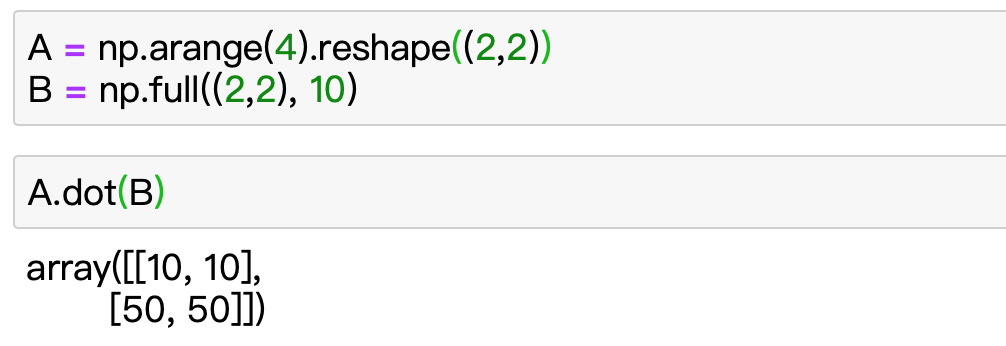
**矩阵减法**

****

**矩阵伪乘法**

****

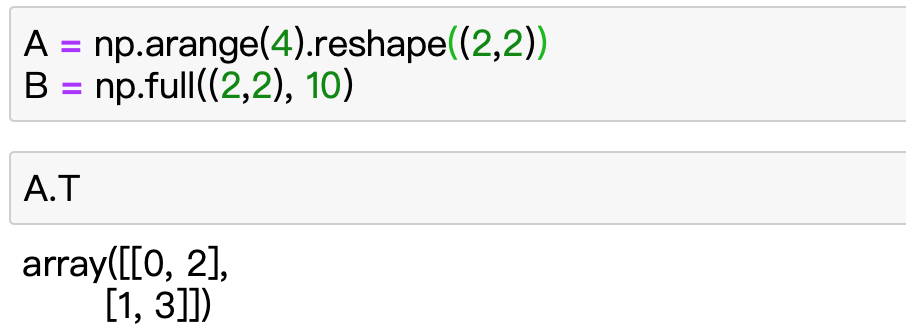
**矩阵乘法**

****

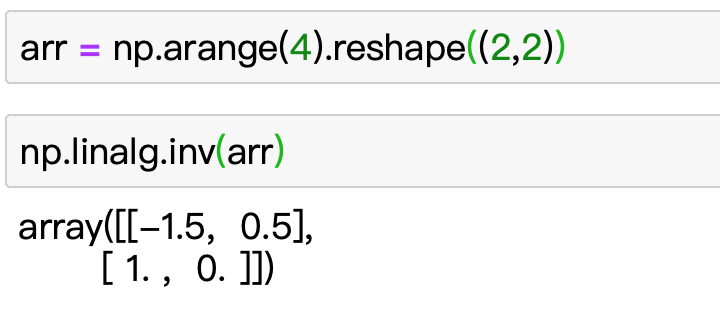
**矩阵除法**

****

**矩阵转置**

****

**矩阵的逆**

****

* Numpy聚合操作

**np.sum()**

**np.min()**

**np.max()**

**np.mean()**

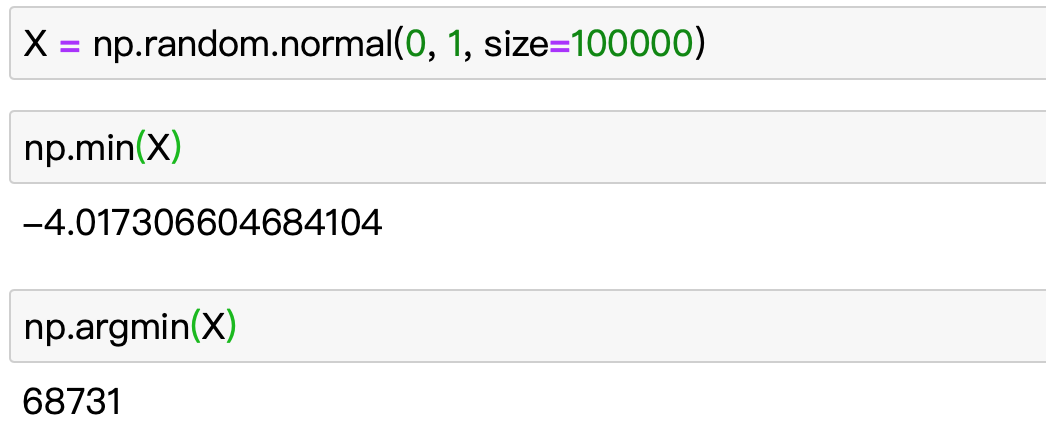
**np.median()**

**np.var()**

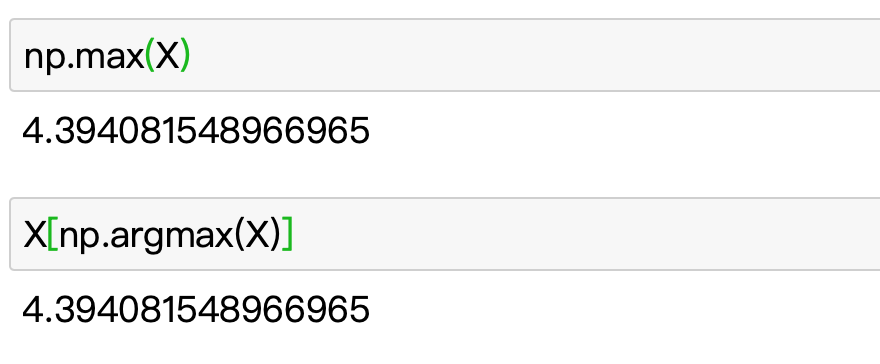
**np.std()**

* Numpy索引操作

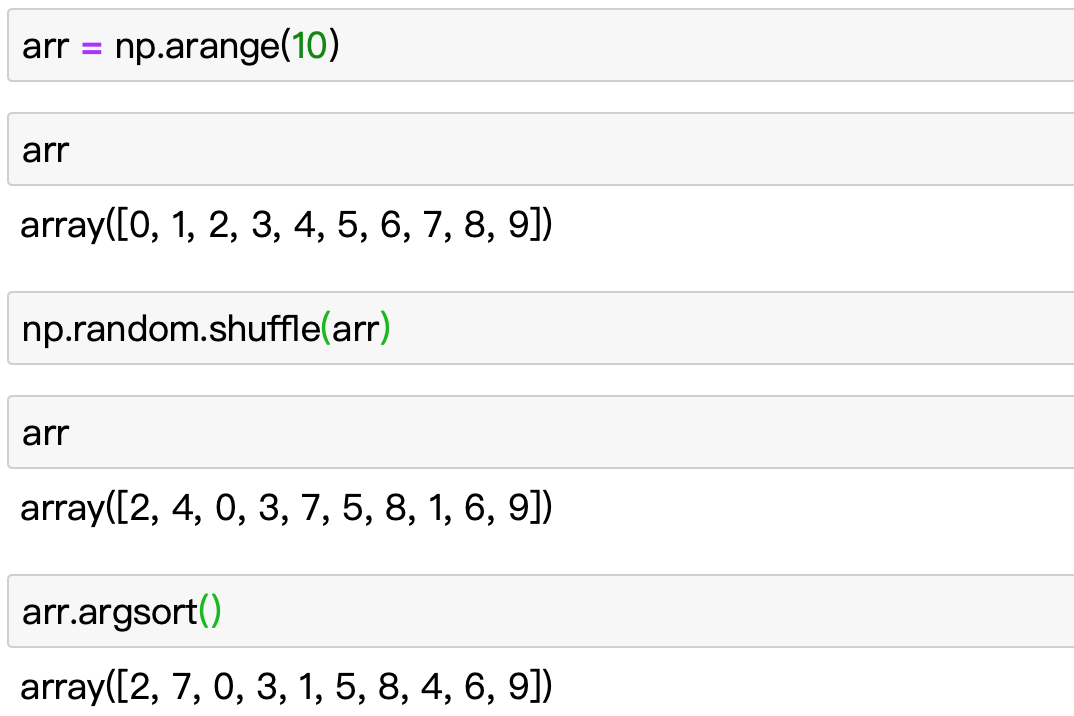
**argmin**

****

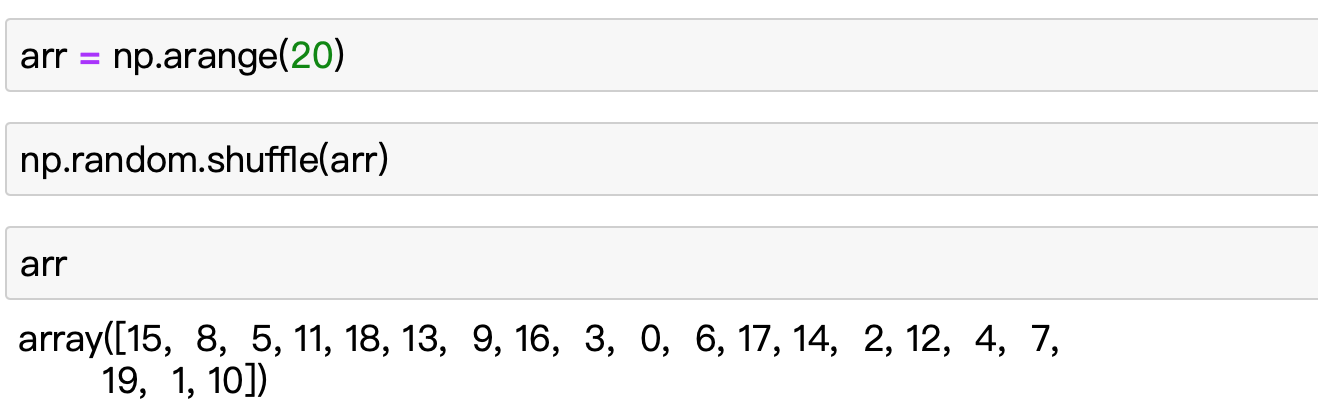
**argmax**



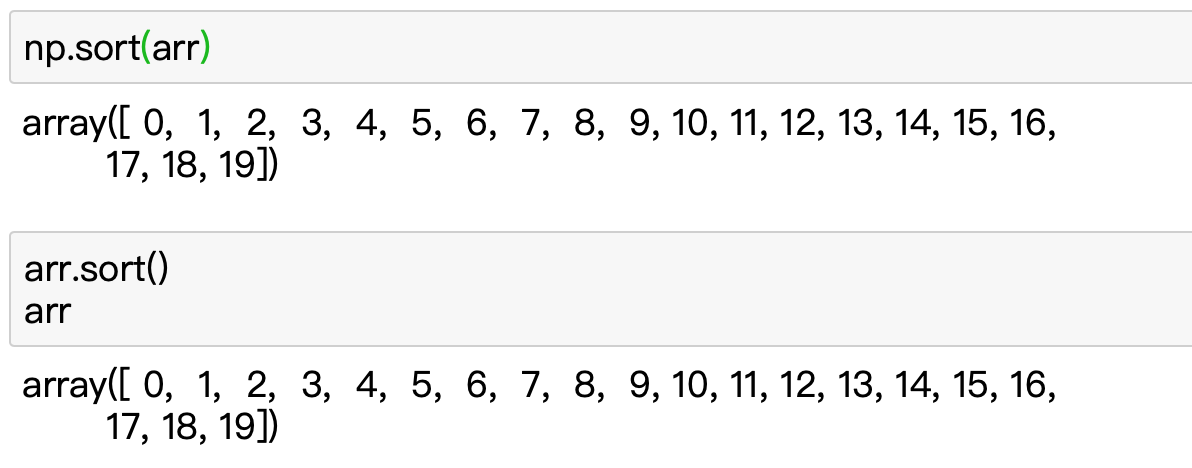
**argsort**

****

**shuffle**

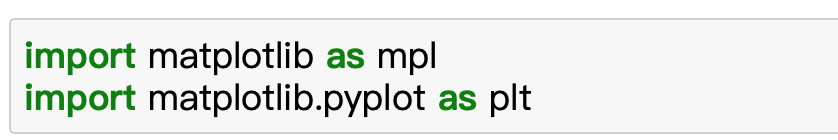


**sort**



* Matplotlib基础

**matplotlib导入**



**直线图**

****

**多类数据绘制**

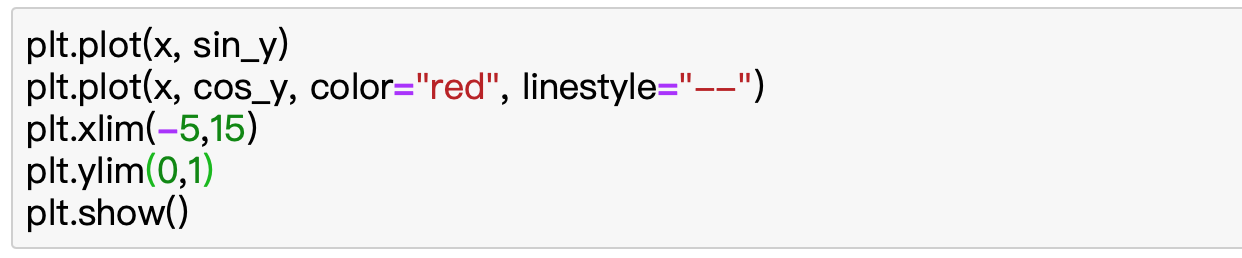


**参数**

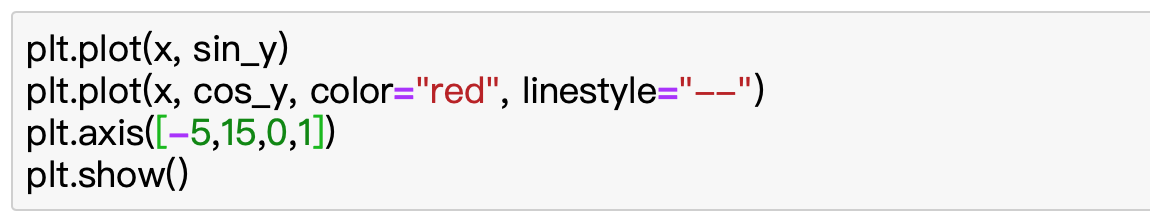
**1.color 颜色**

**2.linestyle 线条样式**

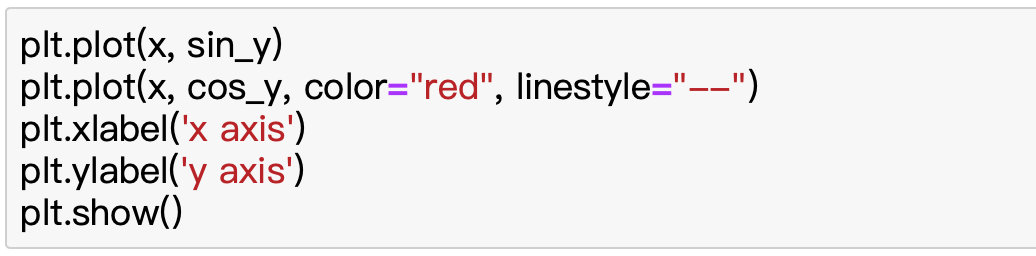
**XY轴值域设置**

****

**同时设置XY轴**

****

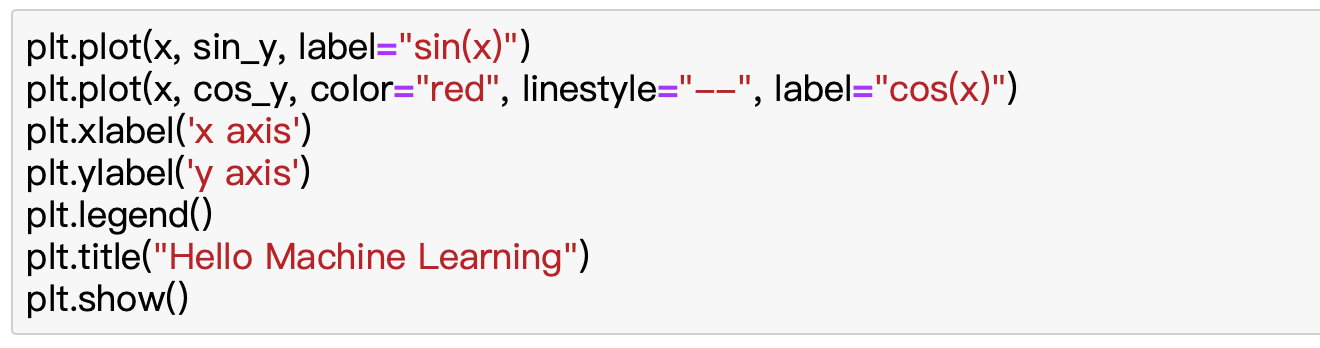
**XY轴标签**

****

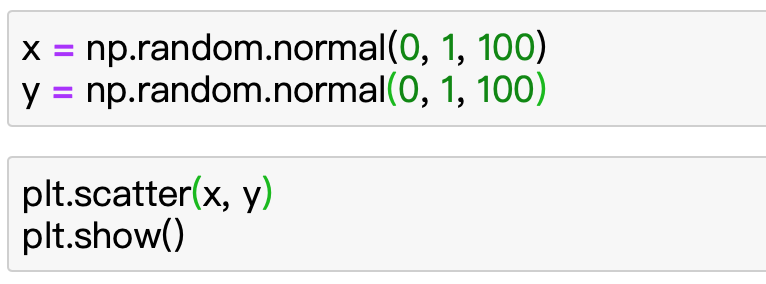
**图形标签**

****

**绘图标题**

****

**散点图**

****

**参数**

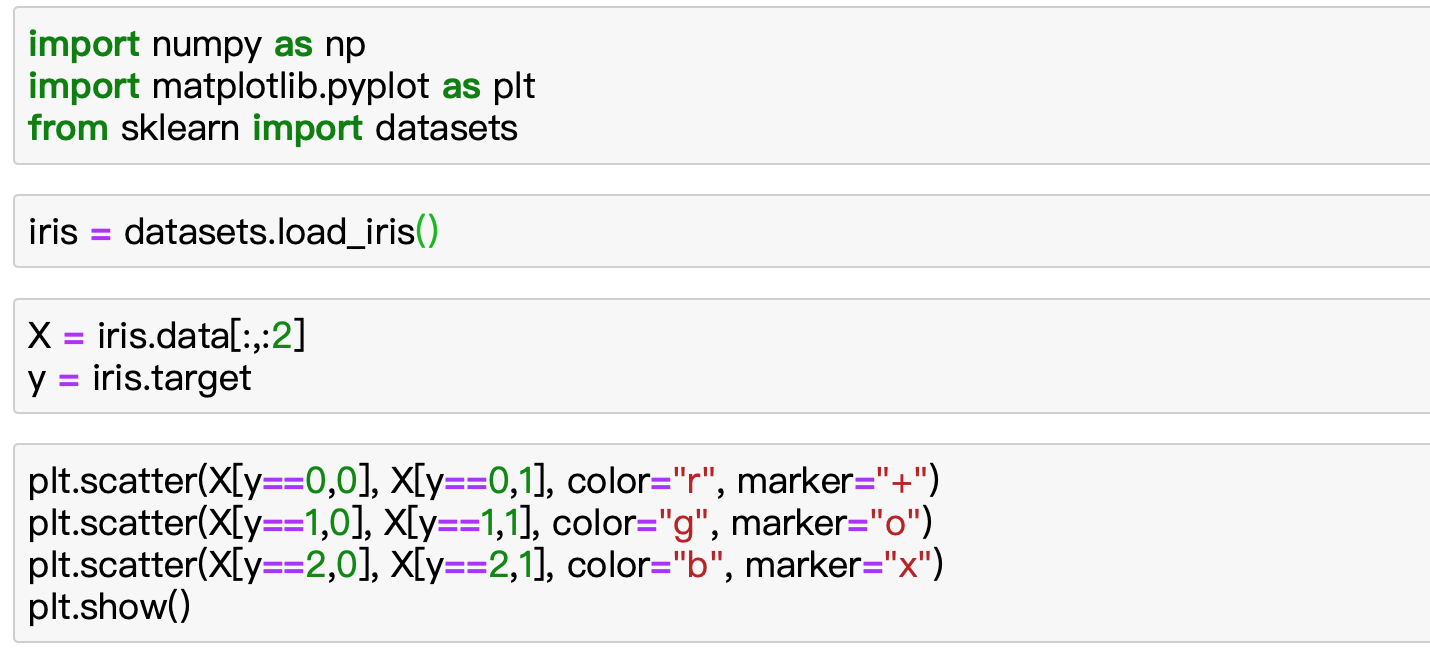
**1.color 颜色**

**2.alpha 透明度**

**3.marker 散点形状**

* Sklearn数据探索

**加载数据集**

****