NumPy 使用 70 问

NumPy 是 Python 科学计算的基础软件包，提供多维数组对象，多种派生对象（掩码数组、矩阵等）以及用于快速操作数组的函数及 API，它包括数学、逻辑、数组形状变换、排序、选择、I/O 、离散傅立叶变换、基本线性代数、基本统计运算、随机模拟等等。

1. 将 NumPy 导入为 np，并查看版本

问题：将 NumPy 导入为 np，并输出版本号。

```
import numpy as np
print(np.__version__)
#> 1.13.3
```

2. 如何创建 1 维数组?

问题：创建数字从 0 到 9 的 1 维数组。

期望输出：`#> array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])`

```
import numpy as np

arr = np.arange(10)
print(arr)
```

3. 如何创建 boolean 数组?

问题：创建所有 True 的 3×3 NumPy 数组。

```
import numpy as np

ns1 = np.full((3, 3), True, dtype=bool)
#> array([[ True,  True,  True],
#>        [ True,  True,  True],
#>        [ True,  True,  True]], dtype=bool)
print(ns1)

# Alternate method:
ns = np.ones((3,3), dtype=bool)
print(ns)

```

4. 如何从 1 维数组中提取满足给定条件的项?

问题：从 arr 中提取所有奇数。

输入： >>arr = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])`

期望输出：>>#> array([1, 3, 5, 7, 9])

```
1 import numpy as np
2
3 arr = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
4 ns = arr[arr % 2 == 1]
5 print(ns)
```

5. 如何将 NumPy 数组中满足给定条件的项替换成另一个数值?

问题：将 arr 中的所有奇数替换成 -1。

输入：arr = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])

期望输出：#> array([0, -1, 2, -1, 4, -1, 6, -1, 8, -1])

```
1 import numpy as np
2
3 arr = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
4 arr[arr % 2 == 1] = -1
5 print(arr)
6
```

6. 如何在不影响原始数组的前提下替换满足给定条件的项?

问题：将 arr 中所有奇数替换成 -1，且不改变 arr。

输入：arr = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])

期望输出：out#> array([0, -1, 2, -1, 4, -1, 6, -1, 8, -1])

arr#> array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])

```
1 import numpy as np
2
3 arr = np.arange(10)
4 out = np.where(arr % 2 == 1, -1, arr)
5 print(arr)
6 print(out)
```

7. 如何重塑（reshape）数组?

问题：将 1 维数组转换成 2 维数组（两行）。

输入： `np.arange(10)`

```
array([ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
```

期望输出： `#> array([[0, 1, 2, 3, 4],`

```
#> [5, 6, 7, 8, 9]])
```

```
1 import numpy as np
2
3 arr = np.arange(10)
4 arr.reshape(2, -1)
5 print(arr)
```

8. 如何垂直堆叠两个数组?

问题：垂直堆叠数组 a 和 b。

输入： `a = np.arange(10).reshape(2,-1)b = np.repeat(1, 10).reshape(2,-1)`

期望输出： `#> array([[0, 1, 2, 3, 4],`

```
#> [5, 6, 7, 8, 9],
#> [1, 1, 1, 1, 1],
#> [1, 1, 1, 1, 1]])
```

```
1 import numpy as np
2
3 a = np.arange(10).reshape(2,-1)
4 b = np.repeat(1, 10).reshape(2,-1)
5
6 r1 = np.concatenate([a, b], axis=0)     # 方法一
7 r2 = np.vstack([a, b])                  # 方法二
8 r3 = np.r_[a, b]                        # 方法三
9
10 print(r1)
11 print(r2)
12 print(r3)
13
```

9. 如何水平堆叠两个数组?

问题：水平堆叠数组 a 和 b。

输入： `a = np.arange(10).reshape(2,-1)b = np.repeat(1, 10).reshape(2,-1)`

期望输出：`#> array([[0, 1, 2, 3, 4, 1, 1, 1, 1, 1],`

`#> [5, 6, 7, 8, 9, 1, 1, 1, 1, 1]])`

```
import numpy as np

a = np.arange(10).reshape(2,-1)
b = np.repeat(1, 10).reshape(2,-1)

r1 = np.concatenate([a, b], axis=1)     # 方法一
r2 = np.hstack([a, b])                  # 方法二
r3 = np.c_[a, b]                        # 方法三
print(r1)
print(r2)
print(r3)
```

10. 在不使用硬编码的前提下，如何在 NumPy 中生成自定义序列?

问题：在不使用硬编码的前提下创建以下模式。仅使用 NumPy 函数和以下输入数组 a。

输入：`a = np.array([1,2,3])`

期望输出：`#> array([1, 1, 1, 2, 2, 2, 3, 3, 3, 1, 2, 3, 1, 2, 3, 1, 2, 3])`

```
import numpy as np

a = np.array([1,2,3])
r1 = np.r_[np.repeat(a, 3), np.tile(a, 3)]
print(r1)
```

11. 如何获得两个 Python NumPy 数组中共同的项?

问题：获取数组 a 和 b 中的共同项。

输入：`a = np.array([1,2,3,2,3,4,3,4,5,6])b = np.array([7,2,10,2,7,4,9,4,9,8])`

期望输出：`array([2, 4])`

```
import numpy as np

a = np.array([1,2,3,2,3,4,3,4,5,6])
b = np.array([7,2,10,2,7,4,9,4,9,8])
r1 = np.intersect1d(a,b)
```

```
print(r1)
```

12. 如何从一个数组中移除与另一个数组重复的项?

问题：从数组 a 中移除出现在数组 b 中的所有项。

输入：`a = np.array([1,2,3,4,5])b = np.array([5,6,7,8,9])`

期望输出：`array([1,2,3,4])`

```
import numpy as np

a = np.array([1,2,3,4,5])
b = np.array([5,6,7,8,9])
r1 = np.setdiff1d(a,b)
print(r1)
```

13. 如何获取两个数组匹配元素的位置?

问题：获取数组 a 和 b 中匹配元素的位置。

输入：`a = np.array([1,2,3,2,3,4,3,4,5,6])b = np.array([7,2,10,2,7,4,9,4,9,8])`

期望输出：`#> (array([1, 3, 5, 7]),)`

```
import numpy as np

a = np.array([1,2,3,2,3,4,3,4,5,6])
b = np.array([7,2,10,2,7,4,9,4,9,8])

r1 = np.where(a == b)
print(r1)
```

14. 如何从 NumPy 数组中提取给定范围内的所有数字?

问题：从数组 a 中提取 5 和 10 之间的所有项。

输入：`a = np.arange(15)`

期望输出：`(array([ 5, 6, 7, 8, 9, 10]),)`

```
import numpy as np

```

```
a = np.arange(15)

index = np.where((a >= 5) & (a <= 10))  # 方法一
r1 = a[index]

index = np.where(np.logical_and(a >= 5, a <= 10))  # 方法二
r2 = a[index]

r3 = a[(a >= 5) & (a <= 10)]  # 方法三
print(r1)
print(r2)
print(r3)

```

15. 如何创建一个 Python 函数以对 NumPy 数组执行元素级的操作?

问题：转换函数 maxx，使其从只能对比标量而变为对比两个数组。

输入：

```
1 def maxx(x, y): 2 """Get the maximum of two items""" 3 if x >= y: 4 retur

n x 5 else: 6 return y 7 8 maxx(1, 5) 9 #> 5
```

期望输出：

```
a = np.array([5, 7, 9, 8, 6, 4, 5])
        b = np.array([6, 3, 4, 8, 9, 7, 1])
        pair_max(a, b)
        #> array([ 6., 7., 9., 8., 9., 7., 5.])
```

```
import numpy as np


def maxx(x, y):
    """Get the maximum of two items"""
    if x >= y:
        return x
    else:
        return y


pair_max = np.vectorize(maxx, otypes=[float])

a = np.array([5, 7, 9, 8, 6, 4, 5])
b = np.array([6, 3, 4, 8, 9, 7, 1])

r1 = pair_max(a, b)
print(r1)
```

16. 如何在 2d NumPy 数组中交换两个列?

问题：在数组 arr 中交换列 1 和列 2。

```
arr = np.arange(9).reshape(3,3)arr
```

```
import numpy as np

arr = np.arange(9).reshape(3,3)
r1 = arr[:, [1,0,2]]
print(r1)
```

17. 如何在 2d NumPy 数组中交换两个行?

问题：在数组 arr 中交换行 1 和行 2。

```
arr = np.arange(9).reshape(3,3)arr
```

```
import numpy as np

arr = np.arange(9).reshape(3, 3)
r1 = arr[[1, 0, 2], :]
print(r1)

```

18. 如何反转 2D 数组的所有行?

问题：反转 2D 数组 arr 中的所有行。

```
# Inputarr = np.arange(9).reshape(3,3)
```

```
import numpy as np

arr = np.arange(9).reshape(3,3)
r1 = arr[::-1]
print(r1)
```

19. 如何反转 2D 数组的所有列?

问题：反转 2D 数组 arr 中的所有列。

- `# Inputarr = np.arange(9).reshape(3,3)`

```
import numpy as np

arr = np.arange(9).reshape(3,3)
r1 = arr[:, ::-1]
print(r1)
```

20. 如何创建一个包含 5 和 10 之间随机浮点的 2 维数组?

问题：创建一个形态为 5×3 的 2 维数组，包含 5 和 10 之间的随机十进制小数。

```
import numpy as np

arr = np.arange(9).reshape(3,3)
# Solution Method 1:
rand_arr = np.random.randint(low=5, high=10, size=(5,3)) + np.random.random((5,3))
print("1",rand_arr)

# Solution Method 2:
rand_arr = np.random.uniform(5,10, size=(5,3))
print("2",rand_arr)
```

21. 如何在 Python NumPy 数组中仅输出小数点后三位的数字?

问题：输出或显示 NumPy 数组 rand_arr 中小数点后三位的数字。

输入：`rand_arr = np.random.random((5,3))`

```
import numpy as np

rand_arr = np.random.random((5,3))
# Create the random array
rand_arr = np.random.random([5,3])
# Limit to 3 decimal places
np.set_printoptions(precision=3)
rand_arr[:4]
print(rand_arr)

```

22. 如何通过禁用科学计数法（如 1e10）打印 NumPy 数组?

问题：通过禁用科学计数法（如 1e10）打印 NumPy 数组 rand_arr。

输入`# Create the random arraynp.random.seed(100)`

`rand_arr = np.random.random([3,3])/1e3rand_arr`

期望输出：`#> array([[ 0.000543, 0.000278, 0.000425],`

`#> [ 0.000845, 0.000005, 0.000122],`

`#> [ 0.000671, 0.000826, 0.000137]])`

```

```

```
import numpy as np

np.set_printoptions(suppress=False)
# Create the random array
np.random.seed(100)
rand_arr = np.random.random([3,3])/1e3

print(rand_arr)
```

23. 如何限制 NumPy 数组输出中项的数目?

问题：将 Python NumPy 数组 a 输出的项的数目限制在最多 6 个元素。

输入：`a = np.arange(15)`

期望输出：`#> array([ 0, 1, 2, ..., 12, 13, 14])`

```
mport numpy as np

np.set_printoptions(threshold=6)
a = np.arange(15)

print(a)
```

24. 如何在不截断数组的前提下打印出完整的 NumPy 数组?

问题：在不截断数组的前提下打印出完整的 NumPy 数组 a。

输入：`np.set_printoptions(threshold=6)`

`a = np.arange(15)`

期望输出：`a#> array([ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14])`

25. 如何向 Python NumPy 导入包含数字和文本的数据集，同时保持文本不变?

问题：导入 iris 数据集，保持文本不变。

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris = np.genfromtxt(url, delimiter=',', dtype='object')
names = ('sepallength', 'sepalwidth', 'petallength', 'petalwidth', 'species')

r1 = iris[:3]      # Print the first 3 rows
print(r1)
```

26. 如何从 1 维元组数组中提取特定的列?

问题：从前一个问题导入的 1 维 iris 中提取文本列 species。

输入：`rl = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis_1d = np.genfromtxt(url, delimiter= , , dtype=None)`

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_1d = np.genfromtxt(url, delimiter=',', dtype=None)
print(iris_1d.shape)

species = np.array([row[4] for row in iris_1d])
r1 = species[:5]
print(r1)

```

27. 如何将 1 维元组数组转换成 2 维 NumPy 数组?

问题：忽略 species 文本字段，将 1 维 iris 转换成 2 维数组 iris_2d。

`url = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis_1d = np.genfromtxt(url, delimiter= , , dtype=None)`

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_1d = np.genfromtxt(url, delimiter=',', dtype=None)

iris_2d = np.array([row.tolist()[:4] for row in iris_1d])        #方法一
r1 = iris_2d[:4]

iris_2d = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0,1,2,3])   #方法二
r2 = iris_2d[:4]
print(r1)
print(r2)

```

28. 如何计算 NumPy 数组的平均值、中位数和标准差?

问题：找出 iris sepallength（第一列）的平均值、中位数和标准差。

url = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis = np.genfromtxt(url, delimiter= , , dtype= object)

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris = np.genfromtxt(url, delimiter=',', dtype='object')
sepallength = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0])

mu, med, sd = np.mean(sepallength), np.median(sepallength), np.std(sepallength)
print(mu, med, sd)
```

29. 如何归一化数组，使值的范围在 0 和 1 之间?

问题：创建 iris sepallength 的归一化格式，使其值在 0 到 1 之间。

输入：url = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datasepallength = np.genfromtxt(url, delimiter= , , dtype= float , usecols=[0])

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
sepallength = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0])

Smax, Smin = sepallength.max(), sepallength.min()
S1 = (sepallength - Smin)/(Smax - Smin)
S2 = (sepallength - Smin)/sepallength.ptp()  # Thanks, David Ojeda!
print(S1)
print(S2)
```

30. 如何计算 softmax 分数?

问题：计算 sepallength 的 softmax 分数。

url = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datasepallength = np.genfromtxt(url, delimiter= , , dtype= float , usecols=[0])

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris = np.genfromtxt(url, delimiter=',', dtype='object')
sepallength = np.array([float(row[0]) for row in iris])

def softmax(x):
```

```
    """Compute softmax values for each sets of scores in x.
    https://stackoverflow.com/questions/34968722/how-to-implement-the-softmax-function-in-python"""
    e_x = np.exp(x - np.max(x))
    return e_x / e_x.sum(axis=0)

r1 = softmax(sepallength)
print(r1)
```

31. 如何找到 NumPy 数组的百分数?

问题：找出 iris sepallength（第一列）的第 5 个和第 95 个百分数。

```
url = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datasepallength = np.genfromtxt(url, delimiter= , , dtype= float , usecols=[0])
```

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
sepallength = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0])
r1 = np.percentile(sepallength, q=[5, 95])
print(r1)
```

32. 如何在数组的随机位置插入值?

问题：在 iris_2d 数据集中的 20 个随机位置插入 np.nan 值。

```
# Inputurl = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis_2d = np.genfromtxt(url, delimiter= , , dtype= object )
```

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_2d = np.genfromtxt(url, delimiter=',', dtype='object')

i, j = np.where(iris_2d)
np.random.seed(100)
iris_2d[np.random.choice((i), 20), np.random.choice((j), 20)] = np.nan

# Method 2
np.random.seed(100)
iris_2d[np.random.randint(150, size=20), np.random.randint(4, size=20)] = np.nan

# Print first 10 rows
print(iris_2d[:10])
```

33. 如何在 NumPy 数组中找出缺失值的位置?

问题：在 iris_2d 的 sepallength（第一列）中找出缺失值的数目和位置。

```
# Inputurl = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis_2d = np.genfromtxt(url, delimiter= , , dtype= float )iris_2d[np.random.randint(150, size=20), np.random.randint(4, size=20)] = np.nan
```

```
url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_2d = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0,1,2,3])
iris_2d[np.random.randint(150, size=20), np.random.randint(4, size=20)] = np.nan

print("Number of missing values: \n", np.isnan(iris_2d[:, 0]).sum())
print("Position of missing values: \n", np.where(np.isnan(iris_2d[:, 0])))
```

34. 如何基于两个或以上条件过滤 NumPy 数组?

问题：过滤 iris_2d 中满足 petallength（第三列）> 1.5 和 sepallength（第一列）< 5.0 的行。

```
# Inputurl = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis_2d = np.genfromtxt(url, delimiter= , , dtype= float , usecols=[0,1,2,3])
```

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_2d = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0,1,2,3])

condition = (iris_2d[:, 2] > 1.5) & (iris_2d[:, 0] < 5.0)
rl = iris_2d[condition]
print(rl)
```

35. 如何在 NumPy 数组中删除包含缺失值的行?

问题：选择 iris_2d 中不包含 nan 值的行。

```
# Inputurl = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis_2d = np.genfromtxt(url, delimiter= , , dtype= float , usecols=[0,1,2,3])
```

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_2d = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0,1,2,3])
iris_2d[np.random.randint(150, size=20), np.random.randint(4, size=20)] = np.nan
```

```
6
7 # Method 1:
8 any_nan_in_row = np.array([~np.any(np.isnan(row)) for row in iris_2d])
9 r1 = iris_2d[any_nan_in_row][:5]
10
11 # Method 2: (By Rong)
12 r2 = iris_2d[np.sum(np.isnan(iris_2d), axis = 1) == 0][:5]
13 print(r1)
14 print(r2)
```

36. 如何找出 NumPy 数组中两列之间的关联性?

问题：找出 iris_2d 中 SepalLength（第一列）和 PetalLength（第三列）之间的关联性。

```
# Inputurl = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis_2d = np.genfromtxt(url, delimiter= , , dtype= float , usecols=[0,1,2,3])
```

```
1 import numpy as np
2
3 url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
4 iris = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0,1,2,3])
5
6 r1 = np.corrcoef(iris[:, 0], iris[:, 2])[0, 1]
7 print(r1)
8
```

37. 如何确定给定数组是否有空值?

问题：确定 iris_2d 是否有缺失值。

```
# Inputurl = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis_2d = np.genfromtxt(url, delimiter= , , dtype= float , usecols=[0,1,2,3])
```

```
1 import numpy as np
2 url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
3 iris_2d = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0,1,2,3])
4
5 r1 = np.isnan(iris_2d).any()
6 print(r1)
```

38. 如何在 NumPy 数组中将所有缺失值替换成 0?

问题：在 NumPy 数组中将所有 nan 替换成 0。

```
# Inputurl = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis_2d = np.genfromtxt(url, delimiter= , , dtype= float , usecols=[0,1,2,3])iris_2d[np.random.randint(150, size=20), np.random.randint(4, size=20)] = np.nan
```

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_2d = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0,1,2,3])
iris_2d[np.random.randint(150, size=20), np.random.randint(4, size=20)] = np.nan
iris_2d[np.isnan(iris_2d)] = 0
r1 = iris_2d[:4]
print(r1)
```

39. 如何在 NumPy 数组中找出唯一值的数量?

问题：在 iris 的 species 列中找出唯一值及其数量。

```
# Inputurl = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis = np.genfromtxt(url, delimiter= , , dtype= object )names = ( sepallength , sepalwidth , petallength , petalwidth , species )
```

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris = np.genfromtxt(url, delimiter=',', dtype='object')
names = ('sepallength', 'sepalwidth', 'petallength', 'petalwidth', 'species')
species = np.array([row.tolist()[4] for row in iris])
r1 = np.unique(species, return_counts=True)
print(r1)
```

40. 如何将一个数值转换为一个类别（文本）数组?

问题：将 iris_2d 的 petallength（第三列）转换以构建一个文本数组，按如下规则进行转换：

- Less than 3 –> 'small'
- 3-5 –> medium
- >=5 –> large

- ```
  # Inputurl = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis = np.genfromtxt(url, delimiter= , , dtype= object )names = ( sepallength , sepalwidth , petallength , petalwidth , species )
  ```
  ```

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris = np.genfromtxt(url, delimiter=',', dtype='object')
names = ('sepallength', 'sepalwidth', 'petallength', 'petalwidth', 'species')

petal_length_bin = np.digitize(iris[:, 2].astype('float'), [0, 3, 5, 10])
label_map = {1: 'small', 2: 'medium', 3: 'large', 4: np.nan}
petal_length_cat = [label_map[x] for x in petal_length_bin]
r = petal_length_cat[:4]
print(r)
```

41. 如何基于 NumPy 数组现有列创建一个新的列?

问题: 为 iris_2d 中的 volume 列创建一个新的列, volume 指 (pi x petallength x sepal_length^2)/3。

```
url = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis_2d = np.genfromtxt(url, delimiter= , , dtype= object )names = ( sepallength , sepalwidth , petallength , petalwidth , species )
```

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_2d = np.genfromtxt(url, delimiter=',', dtype='object')

sepallength = iris_2d[:, 0].astype('float')
petallength = iris_2d[:, 2].astype('float')
volume = (np.pi * petallength * (sepallength**2))/3
volume = volume[:, np.newaxis]
out = np.hstack([iris_2d, volume])

r1 = out[:4]
print(r1)
```

42. 如何在 NumPy 中执行概率采样?

问题: 随机采样 iris 数据集中的 species 列, 使得 setose 的数量是 versicolor 和 virginica 数量的两倍。

```
# Import iris keeping the text column intacturl = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis = np.genfromtxt(url, delimiter= , , dtype= object )
```

```
import numpy as np
```

```
 2
 3 url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
 4 iris = np.genfromtxt(url, delimiter=',', dtype='object')
 5
 6 species = iris[:, 4]
 7 np.random.seed(100)
 8 a = np.array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'])
 9 species_out = np.random.choice(a, 150, p=[0.5, 0.25, 0.25])
10 np.random.seed(100)
11 probs = np.r_[np.linspace(0, 0.500, num=50), np.linspace(0.501, .750, num=50), np.l
inspace(.751, 1.0, num=50)]
12 index = np.searchsorted(probs, np.random.random(150))
13 species_out = species[index]
14 print(np.unique(species_out, return_counts=True))
15
```

43. 如何在多维数组中找到一维的第二最大值?

问题：在 species setosa 的 petallength 列中找到第二最大值。

```
# Inputurl = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis = np.ge
nfromtxt(url, delimiter= , , dtype= object )names = ( sepallength , sepalwidth , petallength , peta
lwidth , species )
```

```
 1 import numpy as np
 2
 3 url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
 4 iris = np.genfromtxt(url, delimiter=',', dtype='object')
 5
 6 petal_len_setosa = iris[iris[:, 4] == b'Iris-setosa', [2]].astype('float')
 7
 8 r1 = np.unique(np.sort(petal_len_setosa))[-2]
 9 print(r1)
```

44. 如何用给定列将 2 维数组排序?

问题：基于 sepallength 列将 iris 数据集排序。

```
url = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis = np.genfromtx
t(url, delimiter= , , dtype= object )names = ( sepallength , sepalwidth , petallength , petalwidth
 , species )
```

```
 1 import numpy as np
 2
 3 url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
 4 iris = np.genfromtxt(url, delimiter=',', dtype='object')
```

```
names = ('sepallength', 'sepalwidth', 'petallength', 'petalwidth', 'species')

print(iris[iris[:,0].argsort()][:20])
```

45. 如何在 NumPy 数组中找到最频繁出现的值?

问题：在 iris 数据集中找到 petallength（第三列）中最频繁出现的值。

```
url = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis = np.genfromtxt(url, delimiter= , , dtype= object )names = ( sepallength , sepalwidth , petallength , petalwidth , species )
```

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris = np.genfromtxt(url, delimiter=',', dtype='object')
vals, counts = np.unique(iris[:, 2], return_counts=True)
print(vals[np.argmax(counts)])

```

46. 如何找到第一个大于给定值的数的位置?

问题：在 iris 数据集的 petalwidth（第四列）中找到第一个值大于 1.0 的数的位置。

```
# Input:url = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis = np.genfromtxt(url, delimiter= , , dtype= object )
```

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris = np.genfromtxt(url, delimiter=',', dtype='object')

r1 = np.argwhere(iris[:, 3].astype(float) > 1.0)[0]
print(r1)

```

47. 如何将数组中所有大于给定值的数替换为给定的 cutoff 值?

问题：对于数组 a，将所有大于 30 的值替换为 30，将所有小于 10 的值替换为 10。

输入：`np.random.seed(100)np.random.uniform(1,50, 20)`

```
import numpy as np

```

```
np.set_printoptions(precision=2)
np.random.seed(100)
a = np.random.uniform(1,50, 20)

np.clip(a, a_min=10, a_max=30)

print(np.where(a < 10, 10, np.where(a > 30, 30, a)))
```

48. 如何在 NumPy 数组中找到 top-n 数值的位置?

问题：在给定数组 a 中找到 top-5 最大值的位置。

```
np.random.seed(100)a = np.random.uniform(1,50, 20)
```

```
import numpy as np

np.random.seed(100)
a = np.random.uniform(1,50, 20)

# Solution:
print(a.argsort())      #> [18 7 3 10 15]
# Solution 2:
print(np.argpartition(-a, 5)[:5])       #> [15 10  3  7 18]

r1 = a[a.argsort()][-5:]     # Method 1:
print(r1)

r2 = np.sort(a)[-5:]         # Method 2:
print(r2)

r3 = np.partition(a, kth=-5)[-5:]        # Method 3:
print(r3)

r4 = a[np.argpartition(-a, 5)][:5]       # Method 4:
print(r4)
```

49. 如何逐行计算数组中所有值的数量?

问题：逐行计算唯一值的数量。

输入：

```
np.random.seed(100)arr = np.random.randint(1,11,size=(6, 10))
```

```
arr> array([[ 9, 9, 4, 8, 8, 1, 5, 3, 6, 3],> [ 3, 3, 2, 1, 9, 5, 1, 10, 7, 3],> [ 5, 2, 6, 4, 5, 5, 4, 8, 2, 2],> [ 8, 8, 1, 3, 10, 10, 4, 3, 6, 9],> [ 2, 1, 8, 7, 3, 1, 9, 3, 6, 2],> [ 9, 2, 6, 5, 3, 9, 4, 6, 1, 10]])
```

期望输出：`> [[1, 0, 2, 1, 1, 1, 0, 2, 2, 0],> [2, 1, 3, 0, 1, 0, 1, 0, 1, 1],> [0, 3, 0, 2, 3, 1, 0, 1, 0, 0],> [1, 0, 2, 1, 0, 1, 0, 2, 1, 2],> [2, 2, 2, 0, 0, 1, 1, 1, 1, 0],> [1, 1, 1, 1, 1, 2, 0, 0, 2, 1]]`

输出包含 10 个列，表示从 1 到 10 的数字。这些数值分别代表每一行的计数数量。

例如，Cell(0,2) 中有值 2，这意味着，数字 3 在第一行出现了两次。

```
import numpy as np

np.random.seed(100)
arr = np.random.randint(1,11,size=(6, 10))
print(arr)
```

50. 如何将 array_of_arrays 转换为平面 1 维数组?

问题：将 array_of_arrays 转换为平面线性 1 维数组。

`# Input:arr1 = np.arange(3)arr2 = np.arange(3,7)arr3 = np.arange(7,10)array_of_arrays = np.array([arr1, arr2, arr3])array_of_arrays#> array([array([0, 1, 2]), array([3, 4, 5, 6]), array([7, 8, 9])], dtype=object)`

期望输出：`#> array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])`

```
import numpy as np

arr1 = np.arange(3)
arr2 = np.arange(3,7)
arr3 = np.arange(7,10)

array_of_arrays = np.array([arr1, arr2, arr3])
print('array_of_arrays: ', array_of_arrays)

arr_2d = np.array([a for arr in array_of_arrays for a in arr])

arr_2d = np.concatenate(array_of_arrays)
print(arr_2d)
```

51. 如何为 NumPy 数组生成>

问题：计算>

输入：`np.random.seed(101)arr = np.random.randint(1,4, size=6)arr#> array([2, 3, 2, 2, 2, 1])`

输出：`> array([[ 0., 1., 0.],#> [ 0., 0., 1.],#> [ 0., 1., 0.],#> [ 0., 1., 0.],#> [ 0., 1., 0.],#> [ 1., 0., 0.]])`

```
import numpy as np

np.random.seed(101)
arr = np.random.randint(1,4, size=6)
print(arr)
#> array([2, 3, 2, 2, 2, 1])

def one_hot_encodings(arr):
    uniqs = np.unique(arr)
    out = np.zeros((arr.shape[0], uniqs.shape[0]))
    for i, k in enumerate(arr):
        out[i, k-1] = 1
    return out

r1 = one_hot_encodings(arr)
print("r1",r1)
r2 = (arr[:, None] == np.unique(arr)).view(np.int8)
print("r2",r2)
```

52. 如何创建由类别变量分组确定的一维数值?

问题：创建由类别变量分组的行数。使用以下来自 iris species 的样本作为输入。

输入：`url = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.dataspecies = np.genfromtxt(url, delimiter= , , dtype= str , usecols=4)species_small = np.sort(np.random.choice(species, size=20))species_small#> array([Iris-setosa , Iris-setosa , Iris-setosa , Iris-setosa ,#> Iris-setosa , Iris-setosa , Iris-versicolor , Iris-versicolor ,#> Iris-versicolor , Iris-versicolor , Iris-versicolor ,#> Iris-versicolor , Iris-virginica , Iris-virginica ,#> Iris-virginica , Iris-virginica , Iris-virginica ,#> Iris-virginica , Iris-virginica , Iris-virginica],#> dtype= <U15)`

期望输出：`#> [0, 1, 2, 3, 4, 5, 0, 1, 2, 3, 4, 5, 0, 1, 2, 3, 4, 5, 6, 7]`

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
species = np.genfromtxt(url, delimiter=',', dtype='str', usecols=4)
np.random.seed(100)
species_small = np.sort(np.random.choice(species, size=20))
print(species_small)
print([i for val in np.unique(species_small) for i, grp in enumerate(species_small[species_small==val])])
```

53. 如何基于给定的类别变量创建分组 id?

问题：基于给定的类别变量创建分组 id。使用以下来自 iris species 的样本作为输入。

输入：`url = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.dataspecies = np.genfromtxt(url, delimiter= , , dtype= str , usecols=4)species_small = np.sort(np.random.choice(species, size=20))species_small`

```
#> array([ Iris-setosa , Iris-setosa , Iris-setosa , Iris-setosa ,
#> Iris-setosa , Iris-setosa , Iris-versicolor , Iris-versicolor ,
#> Iris-versicolor , Iris-versicolor , Iris-versicolor ,
#> Iris-versicolor , Iris-virginica , Iris-virginica ,
#> Iris-virginica , Iris-virginica , Iris-virginica ,
#> Iris-virginica , Iris-virginica , Iris-virginica ],
#> dtype= <U15 )
```

期望输出：`#> [0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2]`

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
species = np.genfromtxt(url, delimiter=',', dtype='str', usecols=4)
np.random.seed(100)
species_small = np.sort(np.random.choice(species, size=20))
print(species_small)

output = [np.argwhere(np.unique(species_small) == s).tolist()[0][0] for val in np.unique(species_small) for s in species_small[species_small==val]]

# Solution: For Loop version
output = []
uniqs = np.unique(species_small)

for val in uniqs:  # uniq values in group
    for s in species_small[species_small==val]:  # each element in group
        groupid = np.argwhere(uniqs == s).tolist()[0][0]  # groupid
        output.append(groupid)

print(output)
```

54. 如何使用 NumPy 对数组中的项进行排序?

问题：为给定的数值数组 a 创建排序。

输入：`np.random.seed(10)a = np.random.randint(20, size=10)print(a)#> [ 9 4 15 0 17 16 17 8 9 0]`

期望输出：`[4 2 6 0 8 7 9 3 5 1]`

```
import numpy as np

np.random.seed(10)
a = np.random.randint(20, size=10)
print('Array: ', a)

print(a.argsort().argsort())
print('Array: ', a)
```

55. 如何使用 NumPy 对多维数组中的项进行排序?

问题：给出一个数值数组 a，创建一个形态相同的排序数组。

输入：`np.random.seed(10)a = np.random.randint(20, size=[2,5])print(a)#> [[ 9 4 15 0 17]#> [16 17 8 9 0]]`

期望输出：`#> [[4 2 6 0 8]#> [7 9 3 5 1]]`

```
import numpy as np

np.random.seed(10)
a = np.random.randint(20, size=[2,5])
print(a)

print(a.ravel().argsort().argsort().reshape(a.shape))
```

56. 如何在 2 维 NumPy 数组中找到每一行的最大值?

问题：在给定数组中找到每一行的最大值。

`np.random.seed(100)a = np.random.randint(1,10, [5,3])a#> array([[9, 9, 4],#> [8, 8, 1],#> [5, 3, 6],#> [3, 3, 3],#> [2, 1, 9]])`

```
import numpy as np

np.random.seed(100)
a = np.random.randint(1,10, [5,3])
print("a=",a)

r1 = np.amax(a, axis=1)
print("r1=",r1)
r2 = np.apply_along_axis(np.max, arr=a, axis=1)
print("r2",r2)

```

57. 如何计算 2 维 NumPy 数组每一行的 min-by-max?

问题：给定一个 2 维 NumPy 数组，计算每一行的 min-by-max。

```
np.random.seed(100)a = np.random.randint(1,10, [5,3])a#> array([[9, 9, 4],#> [8, 8, 1],#> [5, 3, 6],#> [3, 3, 3],#> [2, 1, 9]])
```

```
import numpy as np

np.random.seed(100)
a = np.random.randint(1,10, [5,3])
print("a=",a)


r1 = np.apply_along_axis(lambda x: np.min(x)/np.max(x), arr=a, axis=1)
print("r1",r1)

```

58. 如何在 NumPy 数组中找到重复条目?

问题：在给定的 NumPy 数组中找到重复条目（从第二次出现开始），并将其标记为 True。第一次出现的条目需要标记为 False。

```
# Inputnp.random.seed(100)a = np.random.randint(0, 5, 10)print( Array: , a)#> Array: [0 0 3 0 2 4 2 2 2 2]
```

期望输出：`#> [False True False True False False True True True True]`

```
import numpy as np

np.random.seed(100)
a = np.random.randint(0, 5, 10)
out = np.full(a.shape[0], True)
unique_positions = np.unique(a, return_index=True)[1]
out[unique_positions] = False

print(out)

```

59. 如何找到 NumPy 的分组平均值?

问题：在 2 维 NumPy 数组的类别列中找到数值的平均值。

输入 `url = https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.datairis = np.genfromtxt(url, delimiter= , , dtype= object )names = ( sepallength , sepalwidth , petallength , petalwidth , species )`

期望解：`#> [[b Iris-setosa , 3.418],#> [b Iris-versicolor , 2.770],#> [b Iris-virginica , 2.974]]`

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris = np.genfromtxt(url, delimiter=',', dtype='object')
names = ('sepallength', 'sepalwidth', 'petallength', 'petalwidth', 'species')

numeric_column = iris[:, 1].astype('float')  # sepalwidth
grouping_column = iris[:, 4]  # species


[[group_val, numeric_column[grouping_column==group_val].mean()] for group_val in np.unique(grouping_column)]

output = []
for group_val in np.unique(grouping_column):
    output.append([group_val, numeric_column[grouping_column==group_val].mean()])

rl = output
print(rl)
```

60. 如何将 PIL 图像转换成 NumPy 数组?

问题：从以下 URL 中导入图像，并将其转换成 NumPy 数组。

URL = https://upload.wikimedia.org/wikipedia/commons/8/8b/Denali_Mt_McKinley.jpg

```
import numpy as np
from io import BytesIO
from PIL import Image
import PIL, requests

# Import image from URL
URL = 'https://upload.wikimedia.org/wikipedia/commons/8/8b/Denali_Mt_McKinley.jpg'
response = requests.get(URL)

I = Image.open(BytesIO(response.content))       # Read it as Image
I = I.resize([150,150])     # Optionally resize
arr = np.asarray(I)     # Convert to numpy array
```

```
13
14 # Optionaly Convert it back to an image and show
15 im = PIL.Image.fromarray(np.uint8(arr))
16 r1 = Image.Image.show(im)
17 print(r1)
18
```

61. 如何删除 NumPy 数组中所有的缺失值?

问题：从 1 维 NumPy 数组中删除所有的 nan 值。

输入：`np.array([1,2,3,np.nan,5,6,7,np.nan])`

期望输出：`array([ 1., 2., 3., 5., 6., 7.])`

```
1 import numpy as np
2
3 a = np.array([1, 2, 3, np.nan, 5, 6, 7, np.nan])
4 r1 = a[~np.isnan(a)]
5 print(r1)
6
```

62. 如何计算两个数组之间的欧几里得距离?

问题：计算两个数组 a 和 b 之间的欧几里得距离。

输入：`a = np.array([1,2,3,4,5])b = np.array([4,5,6,7,8])`

```
1 import numpy as np
2
3 a = np.array([1,2,3,4,5])
4 b = np.array([4,5,6,7,8])
5 dist = np.linalg.norm(a-b)
6 print(dist)
```

63. 如何在一个 1 维数组中找到所有的局部极大值（peak）?

问题：在 1 维数组 a 中找到所有的 peak，peak 指一个数字比两侧的数字都大。

输入：`a = np.array([1, 3, 7, 1, 2, 6, 0, 1])`

期望输出：`#> array([2, 5])`

```
```

```
import numpy as np

a = np.array([1, 3, 7, 1, 2, 6, 0, 1])
doublediff = np.diff(np.sign(np.diff(a)))
peak_locations = np.where(doublediff == -2)[0] + 1
print(peak_locations)
```

64. 如何从 2 维数组中减去 1 维数组，从 2 维数组的每一行分别减去 1 维数组的每一项?

问题：从 2 维数组 a_2d 中减去 1 维数组 b_1d，即从 a_2d 的每一行分别减去 b_1d 的每一项。

输入：`a_2d = np.array([[3,3,3],[4,4,4],[5,5,5]])b_1d = np.array([1,1,1]`

期望输出：`#> [[2 2 2]#> [2 2 2]#> [2 2 2]]`

```
import numpy as np

a_2d = np.array([[3,3,3],[4,4,4],[5,5,5]])
b_1d = np.array([1,2,3])

print(a_2d - b_1d[:,None])
```

65. 如何在数组中找出某个项的第 n 个重复索引?

问题：找到数组 x 中数字 1 的第 5 个重复索引。

```
x = np.array([1, 2, 1, 1, 3, 4, 3, 1, 1, 2, 1, 1, 2])
```

```
import numpy as np

x = np.array([1, 2, 1, 1, 3, 4, 3, 1, 1, 2, 1, 1, 2])
n = 5

[i for i, v in enumerate(x) if v == 1][n-1]
r1 = np.where(x == 1)[0][n-1]
print(r1)
```

66. 如何将 NumPy 的 datetime64 对象（object）转换为 datetime 的 datetime 对象?

问题：将 NumPy 的 datetime64 对象（object）转换为 datetime 的 datetime 对象。

```
Input: a numpy datetime64 objectdt64 = np.datetime64( 2018-02-25 22:10:10 )
```

```
import numpy as np

dt64 = np.datetime64('2018-02-25 22:10:10')

from datetime import datetime
r1 = dt64.tolist()
print(r1)

r2 = dt64.astype(datetime)
print(r2)
```

67. 如何计算 NumPy 数组的移动平均数?

问题：给定 1 维数组，计算 window size 为 3 的移动平均数。

输入：`np.random.seed(100)Z = np.random.randint(10, size=10)`

```
import numpy as np

def moving_average(a, n=3):
    ret = np.cumsum(a, dtype=float)
    ret[n:] = ret[n:] - ret[:-n]
    return ret[n - 1:] / n

np.random.seed(100)
Z = np.random.randint(10, size=10)
print('array: ', Z)

r1 = moving_average(Z, n=3).round(2)
print("r1=", r1)

r2 = np.convolve(Z, np.ones(3) / 3, mode='valid')
print("r2=", r2)
```

68. 给定起始数字、length 和步长，如何创建一个 NumPy 数组序列?

问题：从 5 开始，创建一个 length 为 10 的 NumPy 数组，相邻数字的差是 3。

```
import numpy as np

length = 10
start = 5
step = 3

def seq(start, length, step):
    end = start + (step*length)
    return np.arange(start, end, step)

r1 = seq(start, length, step)
print(r1)
```

69. 如何在不规则 NumPy 日期序列中填充缺失日期?

问题：给定一个非连续日期序列的数组，通过填充缺失的日期，使其变成连续的日期序列。

输入：`# Inputdates = np.arange(np.datetime64( 2018-02-01 ), np.datetime64( 2018-02-25 ), 2)print (dates)`

`#> [ 2018-02-01 2018-02-03 2018-02-05 2018-02-07 2018-02-09#> 2018-02-11 2018-02-13 2018-02-15 2018-02-17 2018-02-19#> 2018-02-21 2018-02-23 ]`

```
import numpy as np

dates = np.arange(np.datetime64('2018-02-01'), np.datetime64('2018-02-25'), 2)
print("dates=", dates)

filled_in = np.array([np.arange(date, (date+d)) for date, d in zip(dates, np.diff(dates))]).reshape(-1)

output = np.hstack([filled_in, dates[-1]])      # add the last day
print("output=", output)

out = []
for date, d in zip(dates, np.diff(dates)):
    out.append(np.arange(date, (date+d)))

filled_in = np.array(out).reshape(-1)
output = np.hstack([filled_in, dates[-1]])      # add the last day
print("output", output)


```

70. 如何基于给定的 1 维数组创建 strides?

问题：给定 1 维数组 arr，使用 strides 生成一个 2 维矩阵，其中 window length 等于 4，strides 等于 2，例如 [[0,1,2,3], [2,3,4,5], [4,5,6,7]..]。

输入：`arr = np.arange(15)arr#> array([ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14])`

期望输出：`#> [[ 0 1 2 3]#> [ 2 3 4 5]#> [ 4 5 6 7]#> [ 6 7 8 9]#> [ 8 9 10 11]#> [10 11 12 13]]`

```
import numpy as np


def gen_strides(a, stride_len=5, window_len=5):
    n_strides = ((a.size - window_len) // stride_len) + 1
    return np.array([a[s:(s + window_len)] for s in np.arange(0, n_strides * stride_len, stride_len)])

print(gen_strides(np.arange(15), stride_len=2, window_len=4))
```