



数据科学基础 I (Matlab)

— 东北大学 —



数据科学中的数学基础与运算基础



相关数学课程

■ 高等数学

■ 线性代数

■ 概率论

■ 统计分析



数据科学中处理的主要数学对象





标量

>> a=4 %可以是整数

a =
4

>> b=5.2 %可以是小数

b =
5.2000

>> x=-1.3 %可以是负数

x =
-1.3000

>> p=pi %可以是某个常数

p =
3.1416

>> y=inf %可以是无穷大

y =
Inf

>> z='M' %可以是一个字母

z =
'M'

>> n=NaN %可以是非数
(Not a Number)

n =
NaN

>> 3==5 %可以是逻辑值

ans =
logical
0



向量 (矢量)

🔍 一组有序排列的数，通过次序索引，可以确定每个单独的数（元素）。

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$x = [x_1 \ x_2 \ \cdots \ x_n]$$

某个事物的若干种状态
某个对象的若干个属性
某个多项式的若干个系数



数列

- 按照一定次序排列的一列数

$$u_1, u_2, \cdots u_n, \cdots$$

通项

- 数列收敛:

$$\lim_{n \rightarrow \infty} u_n = A$$

$$\lim_{n \rightarrow \infty} \frac{1}{2^n} = 0$$

$$\lim_{n \rightarrow \infty} \frac{n}{n+1} = 1$$



向量

```
>> a=[1,2,3,4,5]    %行向量，元素以逗号分隔
```

```
a =
```

```
    1         2         3         4         5
```

```
>> b=[1;2;3;4;5]    %列向量，元素以分号分隔
```

```
b =
```

```
    1  
    2  
    3  
    4  
    5
```

```
>> c=b'    %向量转置实现行列向量转换
```

```
c =
```

```
    1         2         3         4         5
```



向量

```
>> a=1:5    %以:生成向量
```

```
a =
```

```
    1    2    3    4    5
```

```
>> b=1:2:10  %以:和指定步长生成向量
```

```
b =
```

```
    1    3    5    7    9
```




向量

```
>> a=linspace(1,10) %以 linspace生成向量，默认100个
```

```
a =
```

```
1 至 13 列
```

```
1.0000 1.0909 1.1818 ..... 2.0909
```

```
.....
```

```
92 至 100 列
```

```
9.2727 9.3636 9.4545 ..... 10.0000
```

```
>> b=linspace(1,10,5) %以linspace在区间1~10之间生成5个元素的向量
```

```
b =
```

```
1.0000 3.2500 5.5000 7.7500 10.0000
```

```
>> c=logspace(1,5,3) %以logspace在区间1~5之间生成包含3个对数等分数  
据的行向量
```

```
c =
```

```
10 1000 100000
```



矩阵

🔍 矩阵是一个二维数组，其中每一个元素由两个索引所确定。

$$A \in R^{m \times n} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \dots & \dots & & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

一张 $m \times n$ 个像素组成的图片
一个 M 个用户，每个用户 N 个特征的特征数据集



矩阵

>> A=[1,2,3;3,4,5] %同一行的元素用空格或逗号隔开，每一行可以用回车或分号结束

A=

1	2	3
3	4	5

>> b=A(1,:) %提取矩阵的第1行所有列

b=

1	2	3
---	---	---

当矩阵只有一行时：行向量

>> c=A(:,2) %提取矩阵的所有行第2列

c=

2
4

当矩阵只有一列时：列向量

>> d=A(1,3) %提取矩阵的第1行第3列的元素

d=

3

向量是矩阵的特殊形式!



方阵

🔍 当矩阵的行数和列数相等（同为 n ）时，称为 n 阶方阵。

$$A \in R^{n \times n} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

上三角

副对角线

主对角线

下三角



矩阵

```
>> A=[1,2,3;3,4,5;7,8,9]; %3x3的矩阵
>> L=tril(A)                %提取下三角
L =
     1     0     0
     3     4     0
     7     8     9

>> R=triu(A,-1)             %以第-1条对角线分隔提取上三角
R =
     1     2     3
     3     4     5
     0     8     9

>> v=diag(A)                %提取矩阵的主对角线
v =
     1
     4
     9
```



单位矩阵

- `eye(n)`: 生成 $n \times n$ 阶单位矩阵。
- `eye(m,n)`: 生成 $m \times n$ 的单位矩阵，对角线元素为1，其他为0。
- `eye(size(A))`: 生成一个矩阵 A 大小相同的单位矩阵。

```
>> A=eye(3,4) %生成3*4的单位矩阵
```

```
A =
```

1	0	0	0
0	1	0	0
0	0	1	0



全1矩阵

- `ones(n)` : 生成 $n*n$ 的全1矩阵。
- `ones(m,n)` : 生成 $m*n$ 的全1矩阵。
- `ones(size(A))` : 生成与矩阵A大小相同的全1矩阵。
- `ones(m,n,p,...)` 生成 $m*n*p*...$ 的全1的多维矩阵。

```
>> A=ones(3,4) %生成3*4的全1矩阵
```

```
A =
```

1	1	1	1
1	1	1	1
1	1	1	1



全0矩阵

- `zeros(n)` : 生成 $n*n$ 的全0矩阵。
- `zeros(m,n,:)` : 生成 $m*n$ 的全0矩阵。
- `zeros(size(A))` : 生成与矩阵A大小相同的全0矩阵。
- `zeros (m,n,p,...)` : 生成 $m*n*p*...$ 的全0的多维矩阵。

```
>> A=zeros(3,4) %生成3*4的全0矩阵
```

```
A =
```

0	0	0	0
0	0	0	0
0	0	0	0



随机矩阵

- `A=rand(n)`:生成 $n*n$ 的随机矩阵。
- `A=rand(m,n)`:生成 $m*n$ 的随机矩阵。
- `A=rand(size(A))`:生成与矩阵A大小相同的随机矩阵。
- `A=rand(m,n,p,...)`:生成 $m*n*p*...$ 的随机数多维数组。

```
>> A=rand(3,4) %生成3*4的随机矩阵
```

```
A =
```

0.8147	0.9134	0.2785	0.9649
0.9058	0.6324	0.5469	0.1576
0.1270	0.0975	0.9575	0.9706



随机矩阵

- $Y = \text{randn}(n)$: 生成 $n \times n$ 的服从标准正态分布的随机矩阵。
- $Y = \text{randn}(m, n)$: 生成 $m \times n$ 的服从标准正态分布的随机矩阵。
- $Y = \text{randn}(\text{size}(A))$: 生成与矩阵 A 大小相同的服从标准正态分布的随机矩阵。
- $Y = \text{randn}(m, n, p, \dots)$: 生成 $m \times n \times p \times \dots$ 的服从标准正态分布的随机数多维数组。

```
>> A=rand(3,4) %生成3*4的随机矩阵
```

```
A =
```

0.8147	0.9134	0.2785	0.9649
0.9058	0.6324	0.5469	0.1576
0.1270	0.0975	0.9575	0.9706



张量

张量是基于向量和矩阵的推广，一般而言，超过两维的数组叫做张量。例如，可以将任意一张彩色图片表示成一个三阶张量，三个维度分别是图片的高度、宽度和色彩数据。





张量



彩色图像 (张量)



R分量 (矩阵)



G分量 (矩阵)



B分量 (矩阵)



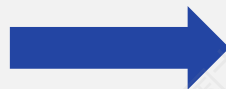
张量

$$x_1 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{pmatrix}$$

$$x_2 = \begin{pmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \\ 7 & 8 & 9 \end{pmatrix}$$

$$x_3 = \begin{pmatrix} 3 & 4 & 5 \\ 4 & 5 & 6 \\ 5 & 6 & 7 \end{pmatrix}$$

张量 $x \in \mathbb{R}^{3 \times 3 \times 3}$



```
>> X=tensor(X)
X is a tensor of size 3 x 3 x 3
X(:,:,1) =
     1     2     3
     3     4     5
     6     7     8
X(:,:,2) =
     2     3     4
     5     6     7
     7     8     9
X(:,:,3) =
     3     4     5
     4     5     6
     5     6     7
```