



# Objectives

- · Hiểu được tầm quan trọng của Khoa học dữ liệu
- Vận dụng được các bước trong quy trình thực hiện một dự án về khoa học dữ liệu.
- Kiến thức, kỹ năng cần thiết để trở thành một nhà khoa học dữ liệu
- Áp dụng được Python và các thư viện phổ biến trong giải quyết một số bài toán cơ bản của Khoa học dữ liệu.

Dr. Ta Quang Chiếu – [E]: <a href="mailto:quangchieu.ta@gmail.com">quangchieu.ta@gmail.com</a> – [M]: 0913 522 275

3/63

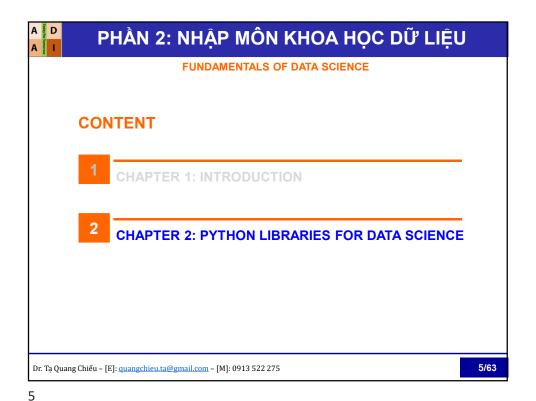
3

## References

- [1]. Introducing Data Science
- [2]. Python Data Science Handbook
- [3]. Data Science from Scratch
- [4]. Python crash course
- [5]. Bài giảng Khoa học dữ liệu

Dr. Tạ Quang Chiều – [E]: <a href="mailto:quangchieu.ta@gmail.com">quangchieu.ta@gmail.com</a> – [M]: 0913 522 275

4/63



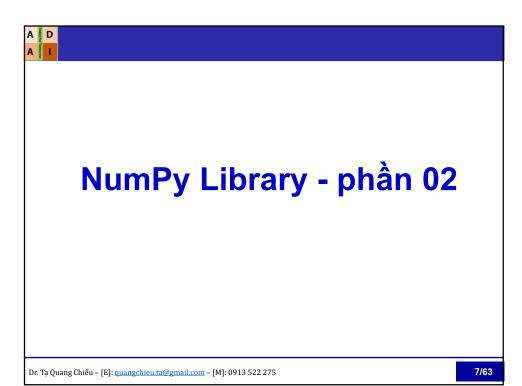
2.1 Python Library for Data Science

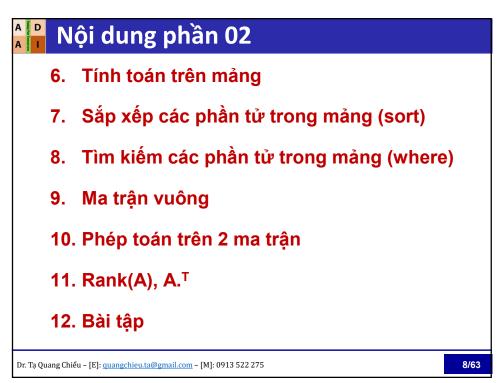
2.2 Numpy library

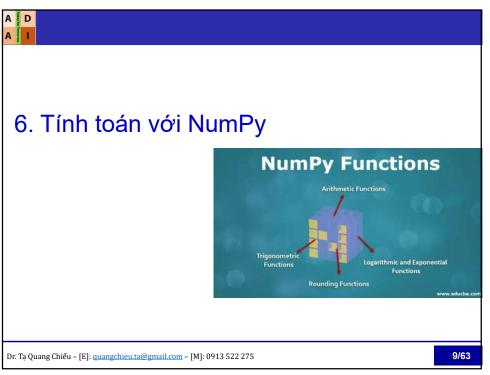
2.3 Pandas library

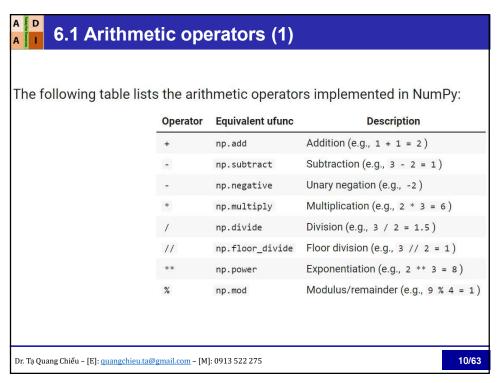
2.4 Matplotlib library

2.5 Scikit-learn library









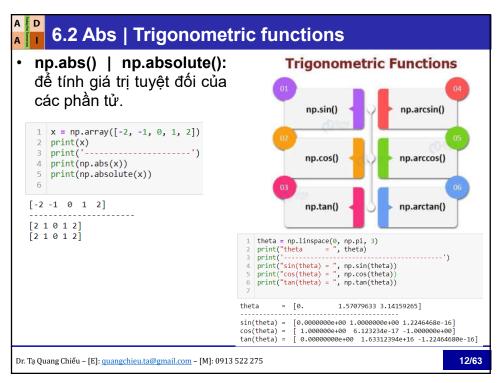
```
6.1 Arithmetic operators (2)
           import numpy as np
         x = np.arange(8)
     3 print("x =", x)
4 print('-----
                                                                                                                          1 print("x
                                                                                                                          #Các phép toán đã giới thiệu trong buổi 01
  5 #Câc phép toán đã giới th
6 print("x + 5 =", x + 5)
7 print("x - 5 =", x - 5)
8 print(" - x = ", -x)
9 print("x * 2 =", x * 2)
10 print("x / 2 =", x // 2)
11 print("x // 2 =", x // 2)
12 print("x % 2 =", x % 2)
13 print("x ^ 3 =", x**3)
                                                                                                                       #Sử dụng các phương thức của NumPy
print("x + 5 = ", np.add(x,5))
print("x - 5 = ", np.subtract(x, 5))
print("x - 2 = ", np.negative(x))
print("x * 2 = ", np.multiply(x,2))
print("x / 2 = ", np.divide(x,2))
print("x / 2 = ", np.floor_divide(x, 2))
print("x % 2 = ", np.mod(x,2))
print("x ^ 3 = ", np.power(x,3))
              = [0 1 2 3 4 5 6 7]
                                                                                                                                      = [0 1 2 3 4 5 6 7]
                                                                                                                      x + 5 = [ 5 6 7 8 9 10 11 12]
 x + 5 = [5 6 7 8 9 10 11 12]
 x - 5 = \begin{bmatrix} -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 \\ -x & = \begin{bmatrix} 0 & -1 & -2 & -3 & -4 & -5 & -6 & -7 \end{bmatrix}
                                                                                                                      x - 5 = \begin{bmatrix} -5 - 4 - 3 - 2 - 1 & 0 & 1 & 2 \\ -x = \begin{bmatrix} 0 - 1 - 2 - 3 - 4 - 5 - 6 - 7 \end{bmatrix}

x * 2 = \begin{bmatrix} 0 & 2 & 4 & 6 & 8 & 10 & 12 & 14 \end{bmatrix}
 x * 2 = [0 2 4 6 8 10 12 14]
 x / 2 = [0. 0.5 1. 1.5 2. 2.5 3. 3.5]

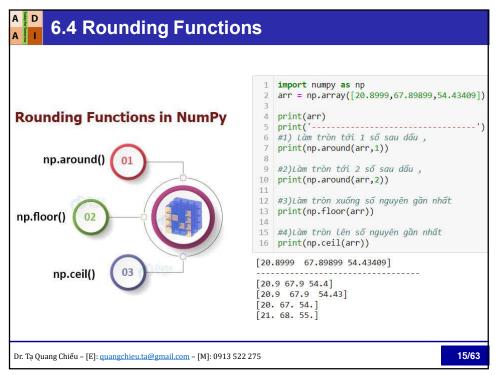
x // 2 = [0 0 1 1 2 2 3 3]

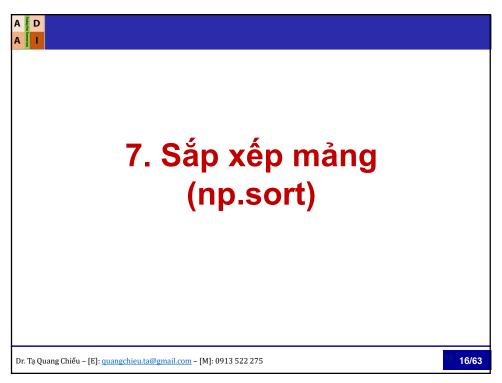
x % 2 = [0 1 0 1 0 1 0 1]
                                                                                                                      x / 2 = [0. 0.5 1. 1.5 2. 2.5 3. 3.5]
x // 2 = [0 0 1 1 2 2 3 3]
                                                                                                                      x \% 2 = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}

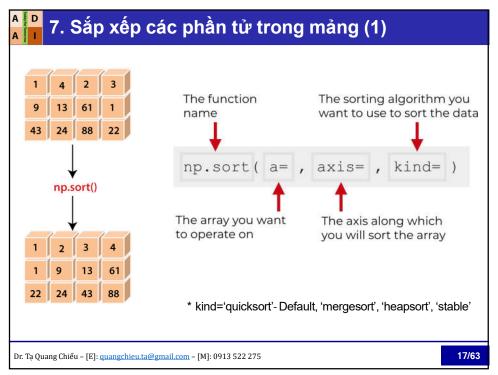
x ^ 3 = \begin{bmatrix} 0 & 1 & 8 & 27 & 64 & 125 & 216 & 343 \end{bmatrix}
                                   1 8 27 64 125 216 343]
                                                                                                                                                                                                              11/63
Dr. Ta Quang Chiếu – [E]: <a href="mailto:quangchieu.ta@gmail.com">quangchieu.ta@gmail.com</a> – [M]: 0913 522 275
```

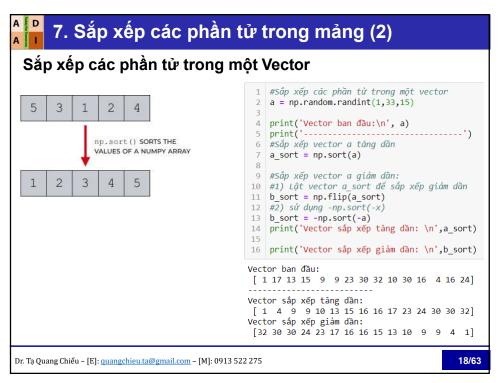


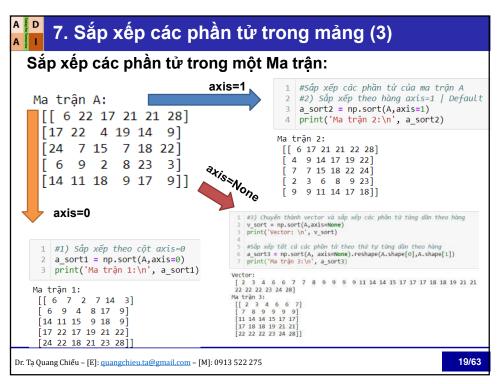
Function	Description
xp(arr)	Returns exponential of an input array element wise
expm1(arr)	Returns exponential exp(x)-1 of an input array element wise
exp2(arr)	Returns exponential 2**x of all elements in an array
og(arr)	Returns natural log of an input array element wise
og10(arr)	Returns log base 10 of an input array element wise
og2(arr)	Returns log base 2 of an input array element wise
ogaddexp(arr)	Returns logarithm of the sum of exponentiations of all inputs
ogaddexp2(arr)	Returns logarithm of the sum of exponentiations of the inputs in base 2

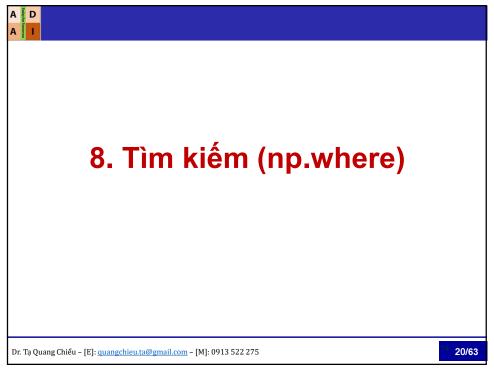


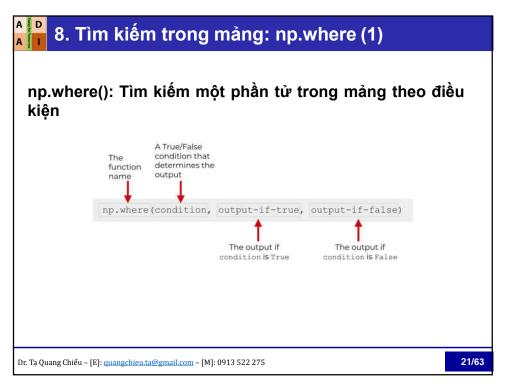






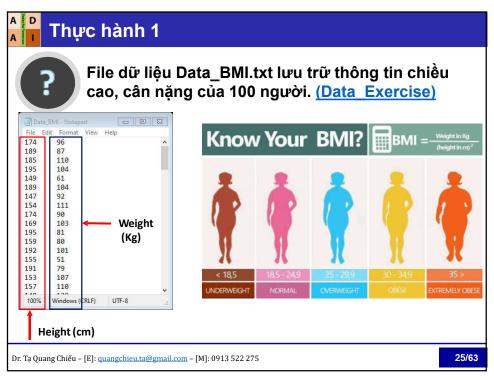


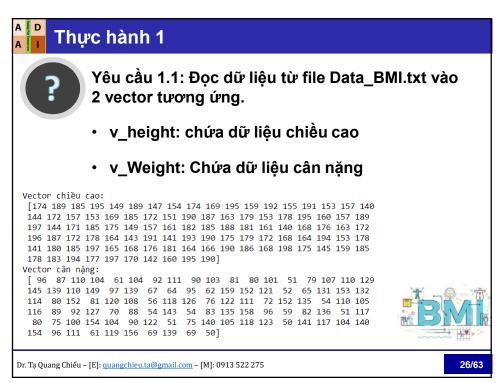




```
**State of the image of the ima
```









### Thực hành 1



- Yêu cầu 1.2: Tạo vector v\_height\_m2 từ vector v height
- theo yêu cầu sau:
  - Chuyển đổi đơn vị của các phần tử từ cm sang m
  - Tính bình phương giá trị các phần tử

Vector v\_height\_m2:

[3.0276 3.5721 3.4225 3.8025 2.2201 3.5721 2.1609 2.3716 3.0276 2.8561 3.8025 2.5281 3.6864 2.4025 3.6481 2.3409 2.4649 1.96 2.0736 2.9584 2.4649 2.3409 2.8561 3.4225 2.9584 2.2801 3.61 3.4969 2.6569 3.2041 2.3409 3.1684 3.8025 2.56 2.4649 3.5721 3.8809 2.0736 2.9241 3.4225 3.0625 2.2201 2.4649 2.5921 3.3124 3.4225 3.5344 3.2761 2.5921 1.96 2.8224 3.0976 2.6569 2.9584 3.8416 3.4969 2.9584 3.1684 2.6896 2.0449 3.6481 1.9881 3.7249 3.61 3.0625 3.2041 2.9584 2.8224 2.6896 3.7636 2.3409 3.1684 1.9881 3.24 3.4225 3.8809 2.7225 2.8224 3.0976 3.2761 2.6896 2.7556 3.61 3.4596 2.8224 3.9204 3.0625 2.1025 2.5281 3.4225 3.1684 3.3489 3.7636 3.1329 3.8809 2.89 2.0164 2.56 3.8025 3.601 ]



Dr. Ta Quang Chiếu – [E]: <a href="mailto:quangchieu.ta@gmail.com">quangchieu.ta@gmail.com</a> – [M]: 0913 522 275

27/63

27



### Thực hành 1



Yêu cầu 1.3: Tính chỉ số BMI của 100 người ngày theo công thức bên dưới, chỉ số BMI được làm tròn tới 1 số sau dấu ",". Lưu kết quả vào vector v\_bmi.

Cân nặng (kg)

Chiều cao x chiều cao (m)

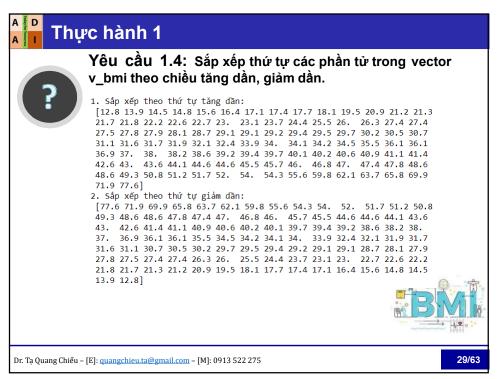
Chỉ số BMI:

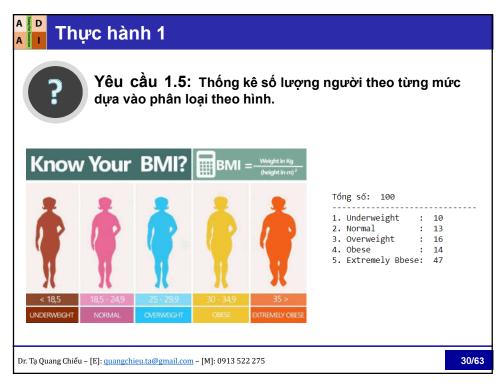
[31.7 24.4 32.1 27.4 27.5 29.1 42.6 46.8 29.7 36.1 21.3 31.6 27.4 21.2 21.7 45.7 44.6 65.8 69.9 47. 44.6 63.7 34. 40.6 22.6 28.1 26.3 17.7 59.8 47.4 51.7 16.4 17.1 51.2 62.1 37. 29.4 38.6 52. 23.7 39.2 48.6 22.7 45.5 38. 22.2 34.5 33.9 27.8 77.6 47.8 17.4 41.4 35.5 30.2 25.5 31.1 40.1 26. 43. 14.8 71.9 14.5 23. 44.1 49.3 32.4 20.9 30.5 36.1 21.8 36.9 40.2 23.1 29.2 39.7 38.2 31.9 39.4 15.6 27.9 50.8 29.1 34.1 43.6 12.8 46. 55.6 41.1 40.9 48.6 28.7 29.5 19.5 30.7 54. 34.2 54.3 18.1 13.9

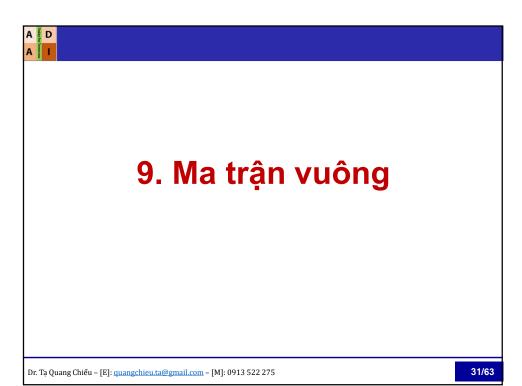


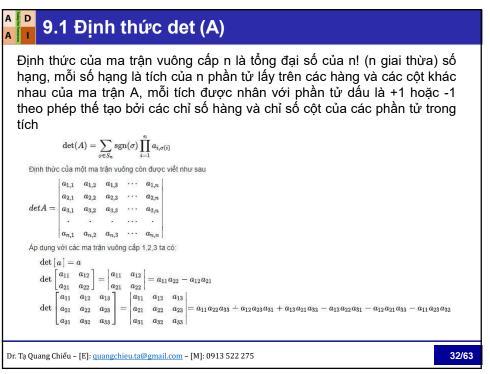
Dr. Tạ Quang Chiểu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

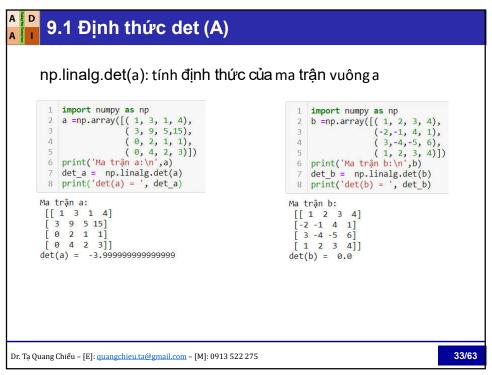
28/63











# 9.2 Ma trận nghịch đảo

Ma trận nghịch đảo của ma trận vuông M ký hiệu M-1

M \* M-1 = I (ma trận đơn vị)

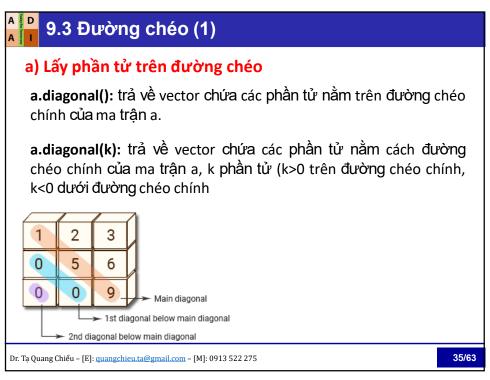
$$M = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{bmatrix}$$

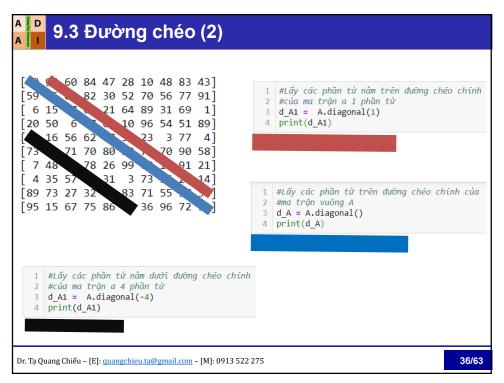
$$M^{-1} = \begin{bmatrix} -24 & 18 & 5 \\ 20 & -15 & -4 \\ -5 & 4 & 1 \end{bmatrix}$$

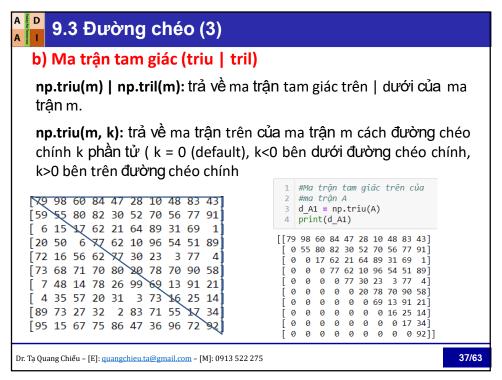
np.linalg.inv(m): Tìm ma trận nghịch đảo của ma trận mdet(m) = 0: Không tồn tại ma trận nghịch đảo

Dr. Tạ Quang Chiểu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

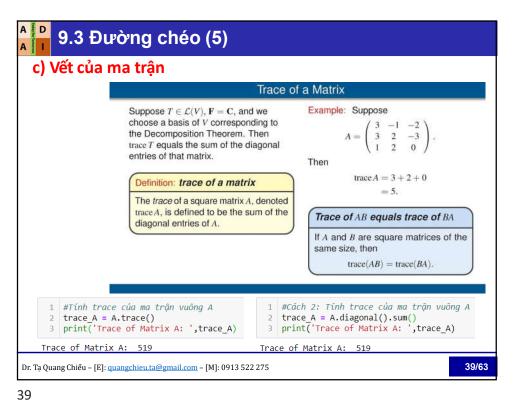
34/63

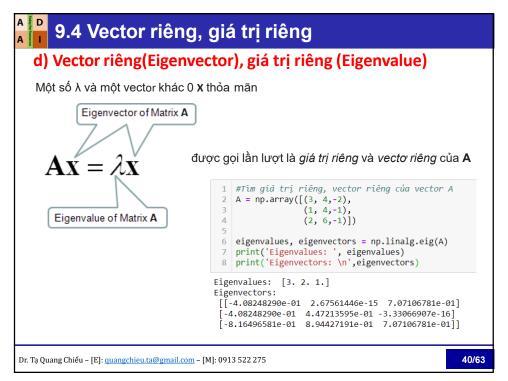


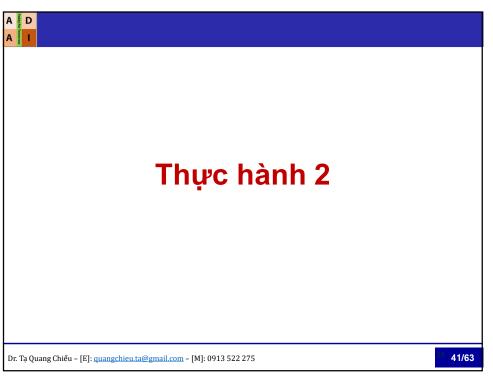


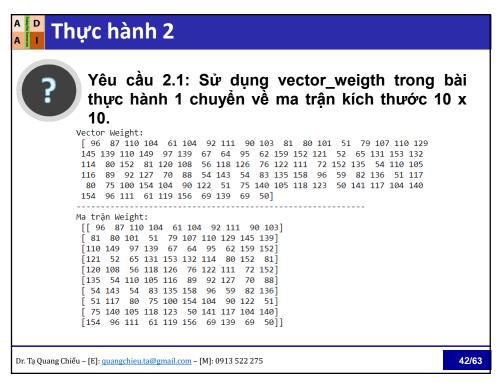


```
9.3 Đường chéo (4)
      #Tạo ma trận là các phần tử trên đường chéo chính
       #của ma trận vuông A cách đường chéo chính
      #về phía trên 2 đường
      d_A1' = np.triu(A,2)
    5 print(d_A1)
         0 60 84 47 28 10 48 83 43]
            0 82 30 52 70 56 77 91]
               0 21 64 89 31 69 1]
               0 10 96 54 51 89]
0 0 23 3 77 4]
                                                1 #Tạo ma trận là các phần tử trên đường chéo chính
     0 0 0 0 0 0
                             0 91 21]
                                                2 #của ma trận vuông A cách đường chéo chính
3 #về phía dưới 3 đường
     0 0 0 0 0 0 0
                             0 14]
     0 0 0 0 0 0 0
                                                4 d_A1 = np.triu(A,-3)
        0 0 0 0 0 0 0 0
                                                5 print(d_A1)
                                               [[13 98 60 84 47 28 10 48 83 43]
[59 5 80 82 30 52 70 56 77 91]
                                                  6 15 7 62 21 64 89 31 69 1
                                                [20 50 6
                                                           62 10 96 54 51 89]
                                                 0 16 56 62 2 30 23 3 77 4]
0 0 71 70 80 2 78 70 90 58]
0 0 0 78 26 99 6 13 91 21]
                                                0 16 56 62
                                                    0 0 0 31 3 73 1 25 14]
0 0 0 0 83 71 55 1 34]
                                                  0
                                                [ 0
                                                     0 0 0 0 0 36 96 72
                                                                                                      38/63
Dr. Ta Quang Chieu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275
```









# Thực hành 2



Yêu cầu 2.2: Cho biết ma trận weight có tồn tại ma trân nghịch đảo không? Nếu có hãy xác định ma trận weight-1?

Yêu cầu 2.3: Tạo vector\_diagonal chứa các phần tử trên đường chéo chính của ma trận weight, tính trace của ma trận weight.

```
a) Đường chéo chính của ma trận weight:
 [ 96 80 97 131 126 89 96 90 104 50]
```

b)Trace của ma trận weight: 959

Dr. Ta Quang Chiếu – [E]: <a href="mailto:quangchieu.ta@gmail.com">quangchieu.ta@gmail.com</a> – [M]: 0913 522 275

43/63

43

## Thực hành 2



Yêu cầu 2.4: Tìm giá trị lớn nhất của ma trận đường chéo trên và ma trận đường chéo dưới không bao gồm các phần tử nằm trên đường chéo chính của ma trận weight?

```
0 87 110 104 61 104 92 111 90 103]
                                                      [[ 0 0
       0 101 51 79 107 110 129 145 139]
0 0 139 67 64 95 62 159 152]
                                                         81 0
                                                                  0
                                                        [110 149 0
                                                                      0 0 0 0
           0 0 153 132 114 80 152 81]
                                                        121 52 65
                                                                      0
                                                                         0 0 0
        0 0 0 0 76 122 111 72 152]
                                                       [120 108 56 118
            0 0 0 0 92 127 70 88]
                                                       [135 54 110 105 116 0
           0 0 0 0 0 59 82 136]
0 0 0 0 0 0 122 51]

    [ 54 143 54 83 135 158 0

    [ 51 117 80 75 100 154 104

       0
              0 0 0 0 0 0 140]
0 0 0 0 0 0 0 0]]
                                                        75 140 105 118 123 50 141 117
                                                       [154 96 111 61 119 156 69 139 69
Phần tử max: 159
                                                      Phần tử max: 158
```

Dr. Ta Quang Chieu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

44/63

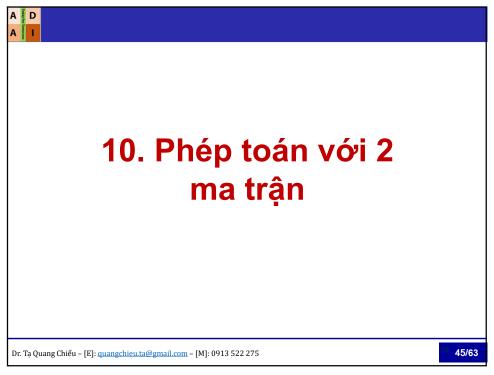
0] 01

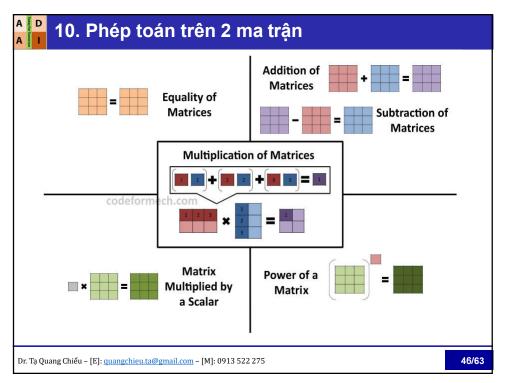
01

0]

01

0





# 10. Phép toán trên 2 ma trận

#### a) So sánh 2 ma trận

np.equal(a,b) | ==: trả về ma trận (T|F) so sánh từng phần tử của ma trận a và ma trận b theo vị trí.

Lưu ý: ma trận a, b có cùng kích thước (m,n)

```
Matrix a:

[[ 9  4  19  1  18]

[15  11  1  9  14]

[17  8  4  10  13]]

Matrix b:

[[ 6  4  9  12  4]

[ 3  6  11  14  10]

[ 1  6  5  12  2]]

[ #1) So sánh 2 ma trận

equal_ab = np.equal(a,b)

#hoặc equal_ab = a==b

print(equal_ab)

[[False True] False False False False
[False False False False False False]]
```

Dr. Ta Quang Chiếu – [E]: <a href="mailto:quangchieu.ta@gmail.com">quangchieu.ta@gmail.com</a> – [M]: 0913 522 275

47/63

47

# A Today For Tempore

## 10. Phép toán trên 2 ma trận

## b) Cộng, trừ 2 ma trận

np.add(a,b) | +: trả về ma trận có các phần tử là tổng của phần tử của ma trận a và ma trận b.

np.suptract(a,b) | -: trả về ma trận có các phần tử là hiệu của phần tử ma trận a và ma trận b

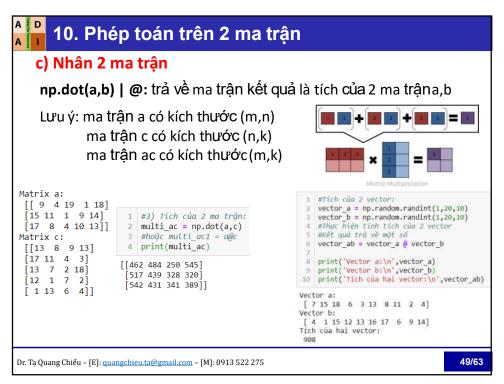
Lưu ý: ma trận a, b có cùng kích thước (m,n)

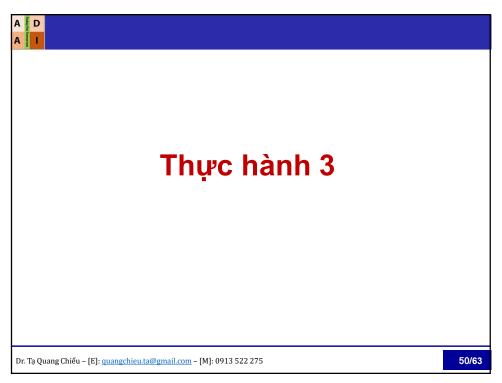
```
1 #Phép cộng 2 ma trận
2 sum_ab = np.add(a,b)
3 #hoặc sum_ab = a + b
4 print (sum_ab)

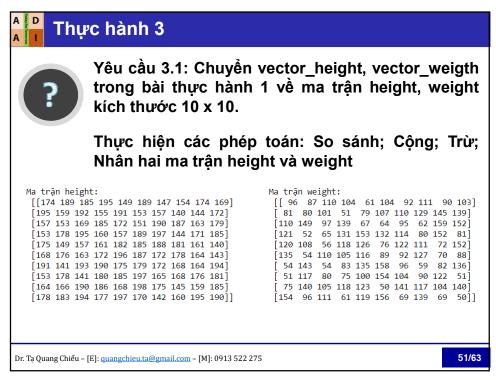
[[15 8 28 13 22]
[18 17 12 23 24]
[18 14 9 22 15]]

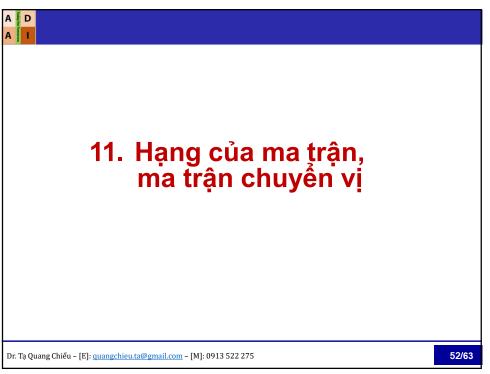
[[16 2 -1 -2 11]]

Dr. Tạ Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275
```









# 11.1 Hạng của ma trận A

Hạng của ma trận là cấp cao nhất của định thức con khác 0 của ma trận đó.

Hạng của ma trận A kí hiệu rank(A) hoặc r(A)

- + Ma trận 0 có hạng bằng 0
- + Ma trận A cấp m x n thì  $0 \le r(A) \le min(m,n)$
- + Ma trận A vuông cấp n:
  - Nếu det(A)  $\neq$  0 thì r(A) = n
  - Nếu det(A) = 0 thì r(A) < n

np.linalg.matrix\_rank(A): Tính hạng của ma trận A

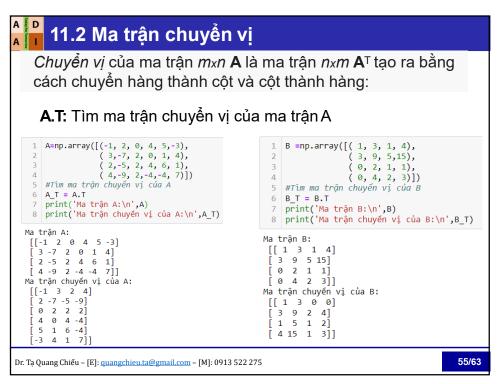
Dr. Ta Quang Chiếu – [E]: quangchieu.ta@gmail.com – [M]: 0913 522 275

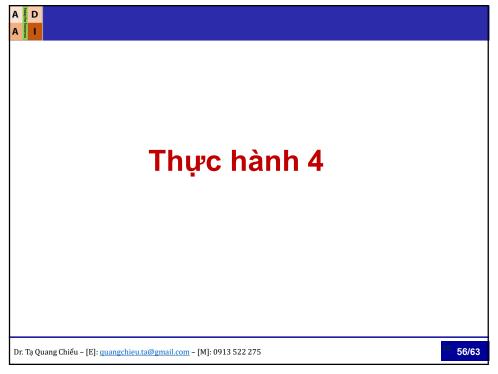
53/63

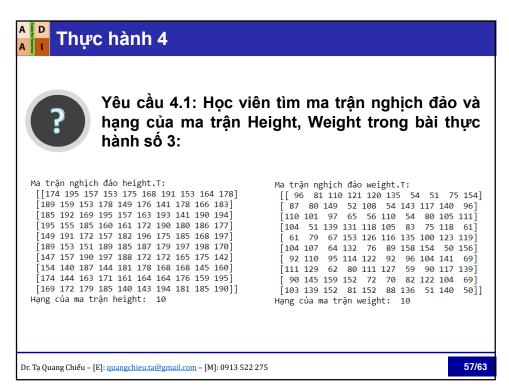
53

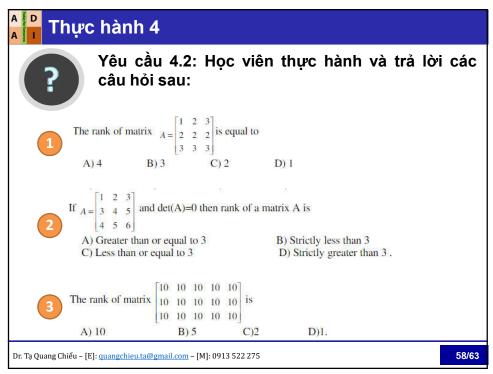
```
11.1 Hạng của ma trận A
  Find the rank and nullity of the matrix
                                                                       A=np.array([(-1, 2, 0, 4, 5,-3),
             \begin{bmatrix} -1 & 2 & 0 & 4 & 5 & -3 \end{bmatrix}
                                                                   A-inp.airay([(1, 2, 0, 4, 3, 3), 3), 3), (3, -7, 2, 0, 1, 4), (2, -5, 2, 4, 6, 1), (4, -9, 2, -4, -4, 7)])

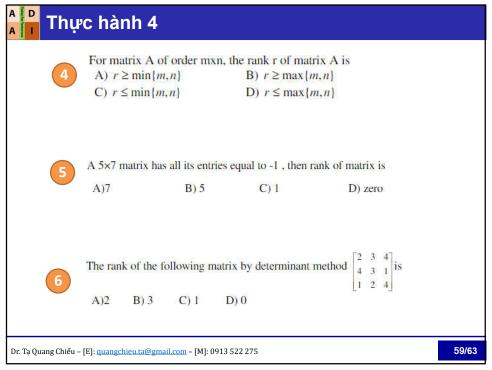
#Tîm hạng của ma trận A
              3 -7 2 0 1 4
             2 -5 2 4 6 1
                                                                   6 rank_a = np.linalg.matrix_rank(A)
 The reduced row-echelon form of A is
                                                                     print(A)
             Γ1 0 -4 -28 -37 13<sup>7</sup>
                                                                   8 print('Rank(A) = ', rank_a)
             0 1 -2 -12 -16 5
             0 0 0 0 0 0
                                                                  [ 3 -7 2 0 1 4]
[ 2 -5 2 4 6 1]
             0 0 0 0 0 0
 Since there are two nonzero rows, the row space and column space
   are both two-dimensional, so rank(A)=2.
  1 #Hạng của ma trận 0
                                                  Ma trận B1:
                                                                                 Ma trận B2:
  A_0 = \text{np.zeros}((4,5))
                                                   [[1 2 3 4]
                                                                                 [[1 3 1 4]
  3 print(A_0)
  rank = np.linalg.matrix_rank(A_0)
print('Rank(A_0) = ', rank)
                                                   [-2 -1 \ 4 \ 1]
                                                                                   [ 3 9 5 15]
                                                   [ 3 -4 -5 6]
                                                                                   [0 2 1 1]
 [[0. 0. 0. 0. 0.]
                                                   [1234]]
                                                                                  [0 4 2 3]]
  [0. 0. 0. 0. 0.]
                                                                                 det(B2) = -4.0
                                                  det(B1) = 0.0
                                                  Rank(B1) = 3
                                                                                 Rank(B2) = 4
  [0. 0. 0. 0. 0.]]
 Rank(A_0) = 0
                                                                                                    54/63
Dr. Ta Quang Chieu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275
```

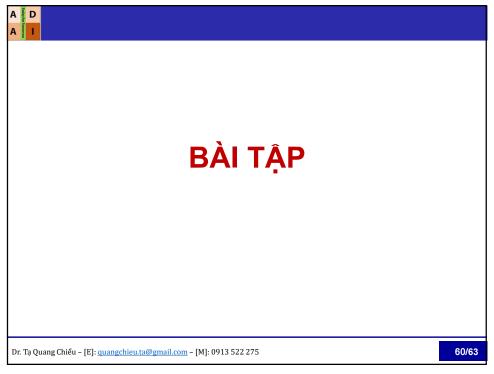












# A A A A

## Bài tập

Bài tập 2.3. Tạo một ma trận 4x4 toàn các giá trị False

Bài tập 2.4. Cho một dãy số nguyên 100 phần tử, hãy tách lấy tất cả những phần tử lẻ cho vào một mảng

Bài tập 2.5. Cho một dãy số tự nhiên 20 phần tử, hãy thay thế tất cả những phần tử lẻ bằng số -1

Bài tập 2.6. Hai mảng a và b có cùng số dòng, hãy ghép chúng theo các dòng thành mảng c, các cột của a rồi đến các cột của b

**Bài tập 2.7**. Mảng a và b có cùng số cột, hãy ghép chúng theo các cột thành mảng c, các dòng của a rồi đến của b

Dr. Ta Quang Chiếu – [E]: <a href="mailto:quangchieu.ta@gmail.com">quangchieu.ta@gmail.com</a> – [M]: 0913 522 275

61/63

61



### Bài tập

**Bài tập 2.8**. Cho một mảng a, hãy in ra tất cả những phần tử trong khoảng từ 5 đến 10

Bài tập 2.9. Sinh ra một mảng số thực có 1000 phần tử, các phần tử nằm trong khoảng từ -0.5 đến < 0.5

Bài tập 2.10. Sinh một ma trận 3x5 gồm các số ngẫu nhiên từ 0 đến nhỏ hơn 10, tính và in ra số lớn nhất trên mỗi dòng của ma trận

Bài tập 2.11. Nhập mảng a và b có 10 phần tử, tính khoảng cách euclid giữa a và b

Dr. Ta Quang Chiếu - [E]: quangchieu.ta@gmail.com - [M]: 0913 522 275

62/63

