

# Machine Learning

Lecturer: Doctor Bui Thanh Hung

Data Science Laboratory

Faculty of Information Technology

Industrial University of Ho Chi Minh city

Email: [hung.buithanhcs@gmail.com](mailto:hung.buithanhcs@gmail.com) ([buithanhhung@iuh.edu.vn](mailto:buithanhhung@iuh.edu.vn))

Website: <https://sites.google.com/site/hungthanhbui1980/>

## Bài 1:

Cho dữ liệu như sau:

Item Number:	1	2	3	4	5
$x$	10	39	19	23	28
$y$	43	13	32	21	20

Sử dụng thư viện tính Ma trận hiệp phương sai của dữ liệu trên

## Bài 2:

Cho dữ liệu như sau:

Item Number:	1	2	3
$x$	1	-1	4
$y$	2	1	3
$z$	1	3	-1

Sử dụng thư viện tính Ma trận hiệp phương sai của dữ liệu trên

## Bài 3:

Hãy tính các Eigenvector and Eigenvalue sử dụng thư viện numpy của các ví dụ sau:

3.1

$$A = \begin{bmatrix} 2 & -4 \\ 3 & -6 \end{bmatrix}$$

3.2

$$A = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}$$

3.3

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

3.4 Viết hàm tính Eigenvector and Eigenvalue tự động từ bất kỳ ma trận đầu vào nào.

#### **Bài 4:**

Giải thuật PCA được trình bày như sau, hãy tự cài đặt lại giải thuật này và áp dụng vào một bộ dữ liệu cụ thể, trực quan kết quả đạt được:

### **Principal Component Analysis**

**Goal:** Find  $r$ -dim projection that best preserves variance

1. Compute mean vector  $\mu$  and covariance matrix  $\Sigma$  of original points
2. Compute eigenvectors and eigenvalues of  $\Sigma$
3. Select top  $r$  eigenvectors
4. Project points onto subspace spanned by them:

$$y = A(x - \mu)$$

where  $y$  is the new point,  $x$  is the old one,  
and the rows of  $A$  are the eigenvectors

#### **Bài 5:**

Cài đặt giải thuật SVD và ứng dụng giải thuật này vào bộ dữ liệu đã dùng ở bài 4.