



Principal Component Analysis (PCA)

Tìm hiểu về thuật toán
PCA

GVHD: Đỗ Văn Tiến

Team Member

Nguyễn Trọng Doanh
19521368

Trương Quốc Bình
19521270



Lưu Anh Dũng
19521392

Table of Contents

01

Dimensionality Reduction

...

02

PCA

...

03

Thuật toán

...

04

Demo code

...

05

Nhận xét

...





01

Dimensionality Reduction

Dimensionality

- **Khái niệm:** Số lượng biến đầu vào hoặc thuộc tính của một bộ data đều được xem như là chiều dữ liệu.



Dimensionality Reduction

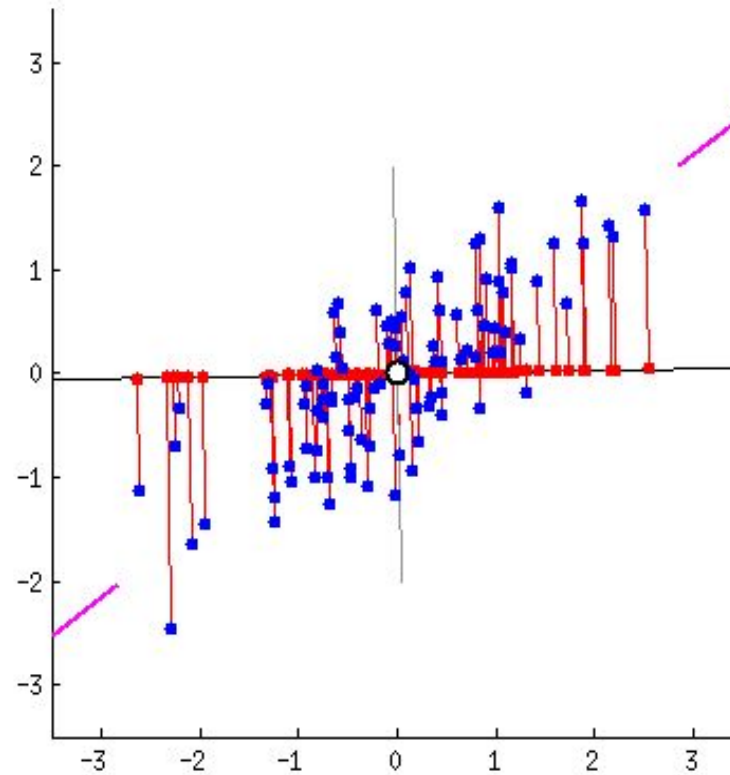


- **Khái niệm:** chuyển đổi dữ liệu từ “nhiều chiều” thành “ít chiều”
 - Dữ liệu “ít chiều” giữ lại một số thuộc tính có ý nghĩa so với dữ liệu gốc

Requirements

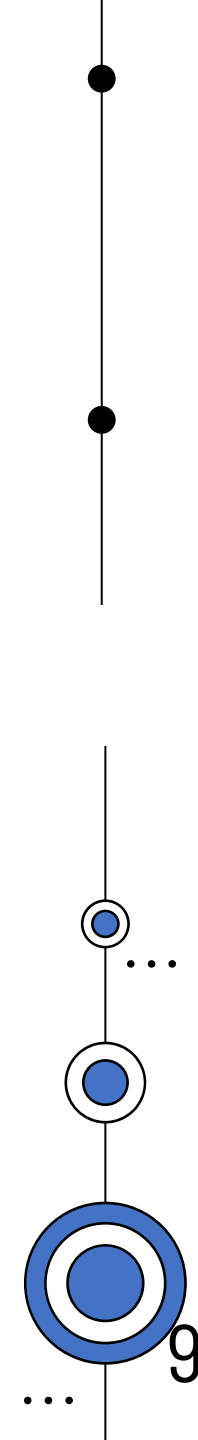
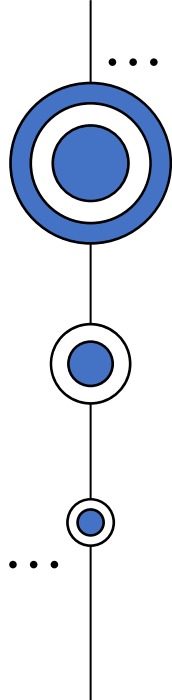
- Dữ liệu ở các chiều phải có **phương sai cao**
- Các chiều dữ liệu phải **độc lập tuyến tính**
- Nếu giảm chiều dữ liệu bằng cách đưa dữ liệu về một hệ trục mới, hệ trục mới phải **giao nhau**

Dimensionality Reduction



02

PCA





Principal Component Analysis



- Khái niệm:

PCA là phương pháp:

- Loại bỏ thành phần phụ
- Giữ thành phần chính
- Không ảnh hưởng nhiều so với dữ liệu gốc

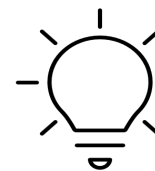
Ứng dụng



...
Nông nghiệp



Kinh tế



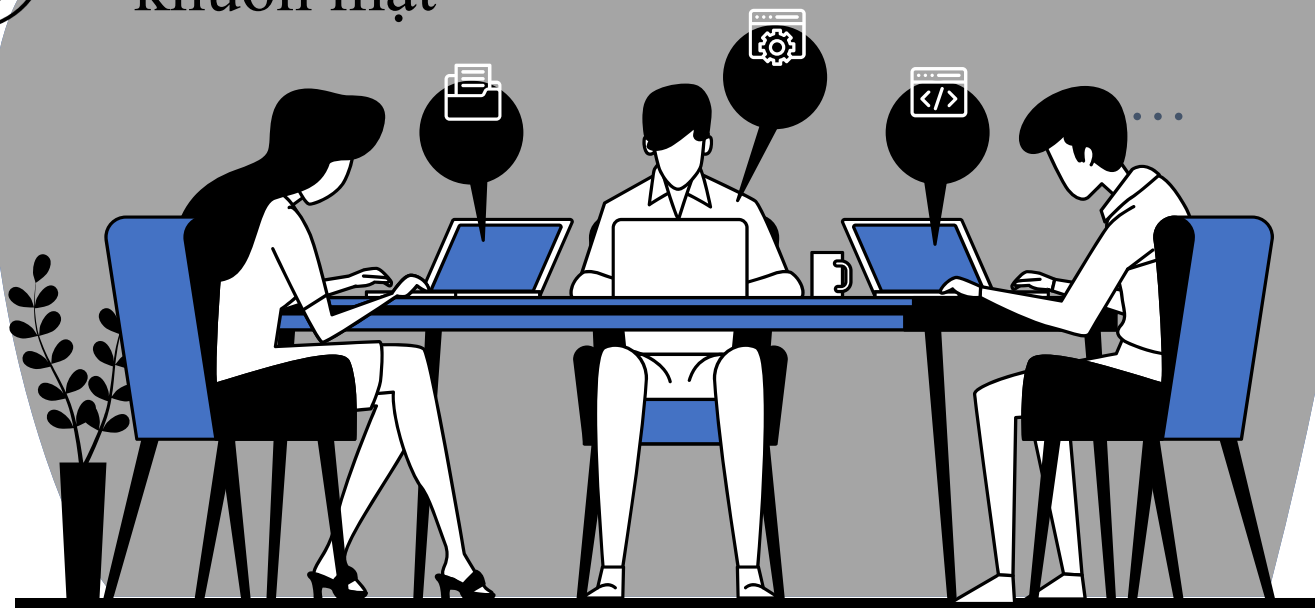
...
Khoa học



Nhận diện
khuôn mặt

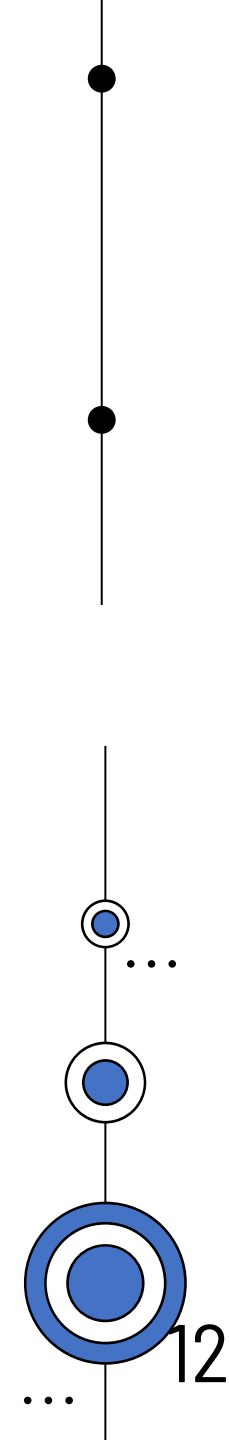
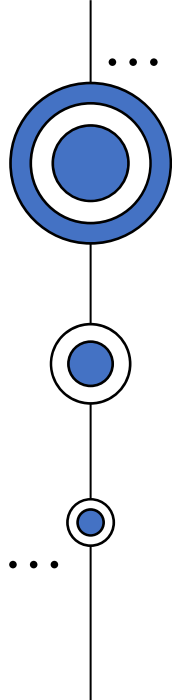


Giảm chiều dữ liệu



03

Thuật toán



1. Tiền xử lý

$$\mathbf{X} = \{x_{ij}\} \in \mathcal{R}^{n \times p}$$

- Centered PC
- Normed PCA

Centered PCA

Mang tất cả các feature (các cột của X) về cùng một gốc tọa độ

$$\hat{\mathbf{X}} = \{\hat{x}_{ij}\}$$

$$\hat{x}_{ij} = \frac{x_{ij} - g_j}{\sqrt{n}}$$

$$g_j = \frac{\sum_{i=1}^n x_{ij}}{n}$$

Normed PCA

Mang tất cả các feature về cùng một gốc tọa độ, đồng thời chuẩn hóa về cùng một quãng standard-deviation bằng 1

$$\hat{\mathbf{X}} = \{\hat{x}_{ij}\}$$

$$\hat{x}_{ij} = \frac{x_{ij} - g_j}{\sqrt{n}\sigma_j}$$

2. Tìm ma trận hiệp phương sai (covariance matrix)

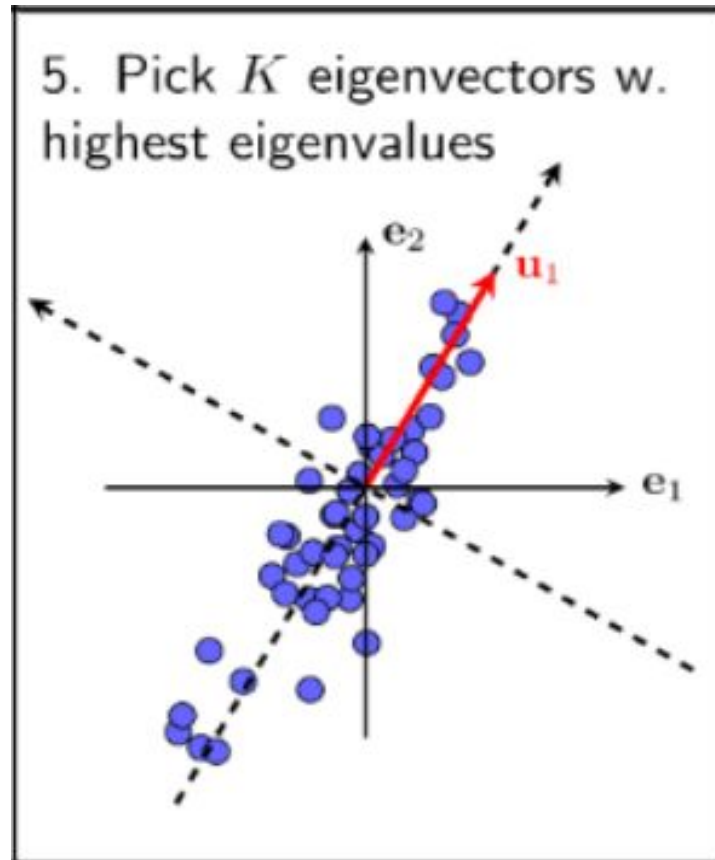
$$S = \frac{1}{N} \hat{X} \hat{X}^T$$

3, 4. Tìm các trị riêng và vector riêng của ma trận

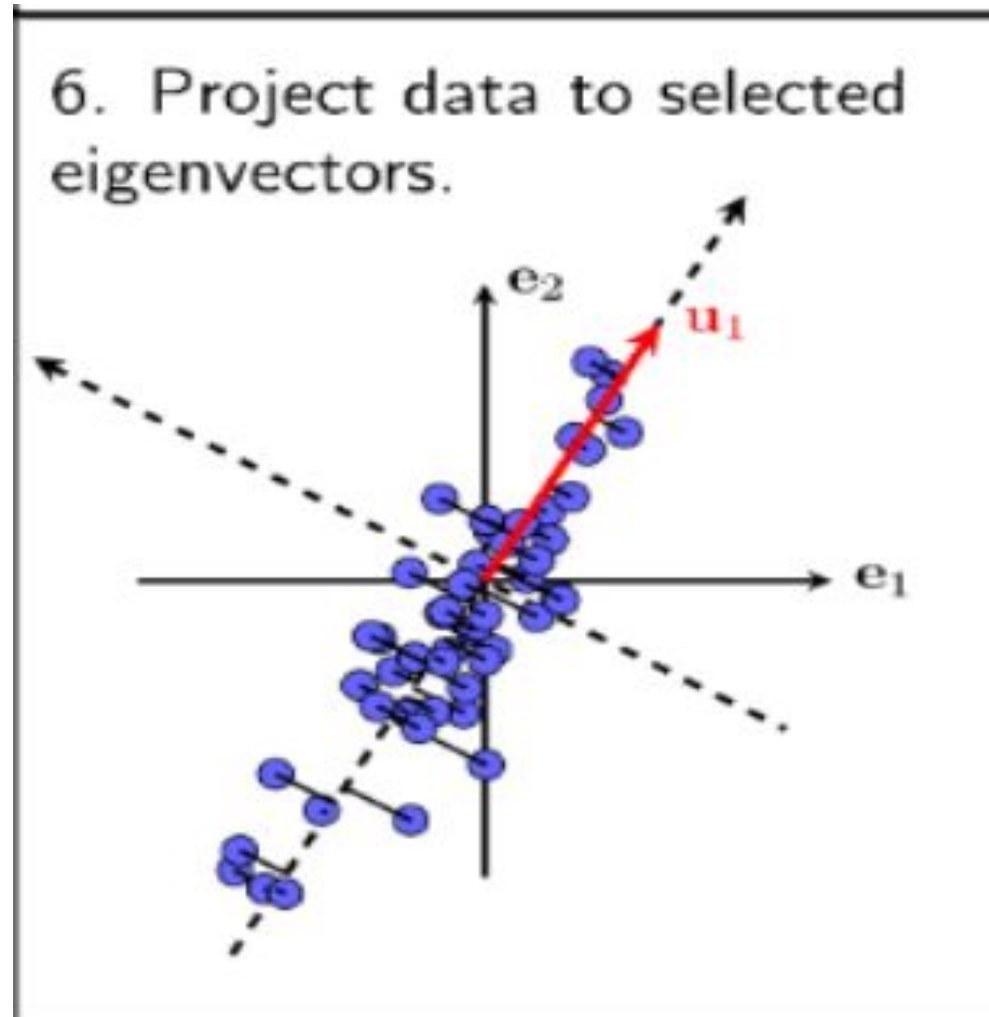
$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$$

$$u_1, u_2, \dots, u_n$$

5. Chọn K vector riêng ứng với K trị riêng lớn nhất để xây dựng ma trận U

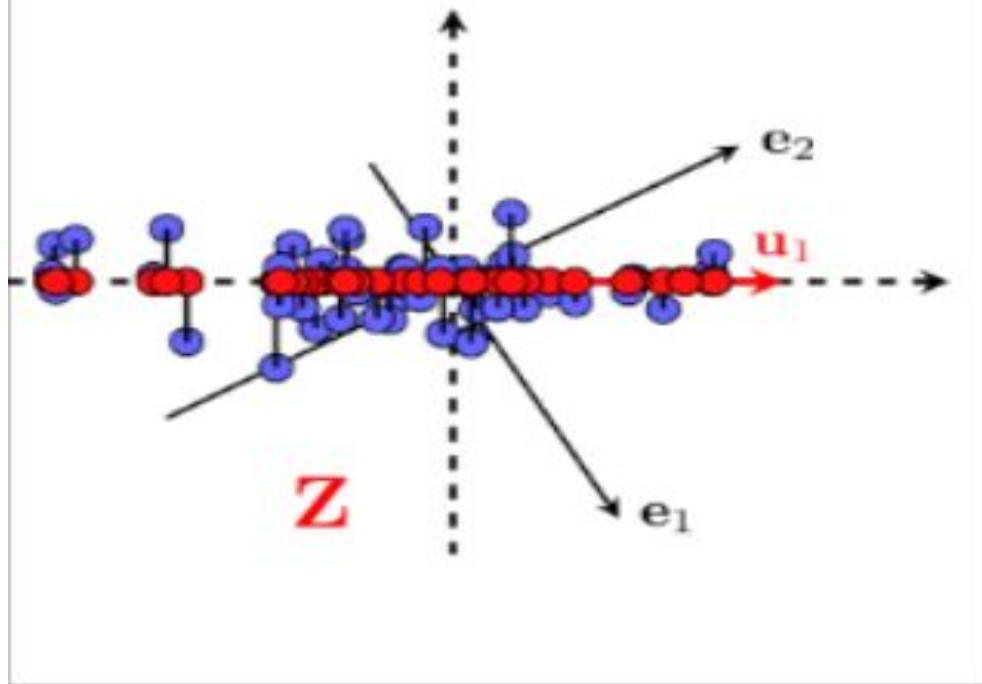


6. Xây dựng ma trận chiều



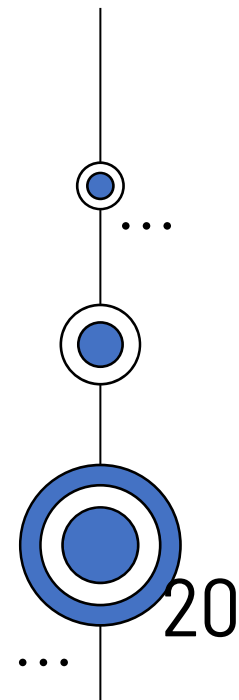
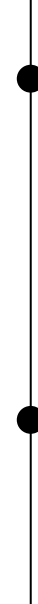
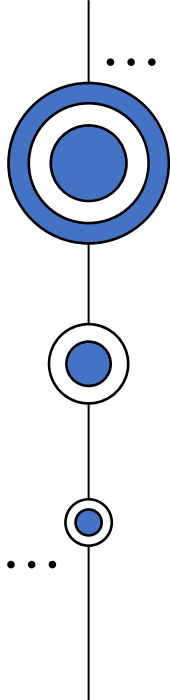
7. Chiếu dữ liệu

7. Obtain projected points in low dimension.



04

Nhận Xét





Ưu điểm

01 Giảm chiều dữ liệu
Nhưng vẫn giữ được chất lượng nhất định so với ban đầu.

02 Nắm bắt thông tin
Dễ quan sát, thu thập nắm bắt thông tin dễ dàng hơn, nhận biết được sự tương quan.

03 Tính toán
Giúp ta tính toán nhanh hơn với độ chính xác cao.



Nhược điểm

01 Các biến độc lập trở nên khó hiểu hơn

02 Một số trường hợp phải chuẩn hóa dữ liệu trước khi sử dụng PCA

03 Mất thông tin

So Sánh

original image



1. ORIGINAL IMAGE



NO OF COLUMNS = 20 FOR COMPRESSED IMAGE

**2. IMAGE COMPRESSED BY
PCA :NO OF COLUMNS = 20**

So Sánh



**3. IMAGE COMPRESSED BY
PCA :NO OF COLUMNS = 35**



**4. IMAGE COMPRESSED BY
JPEG :QUALITY FACTOR=30**

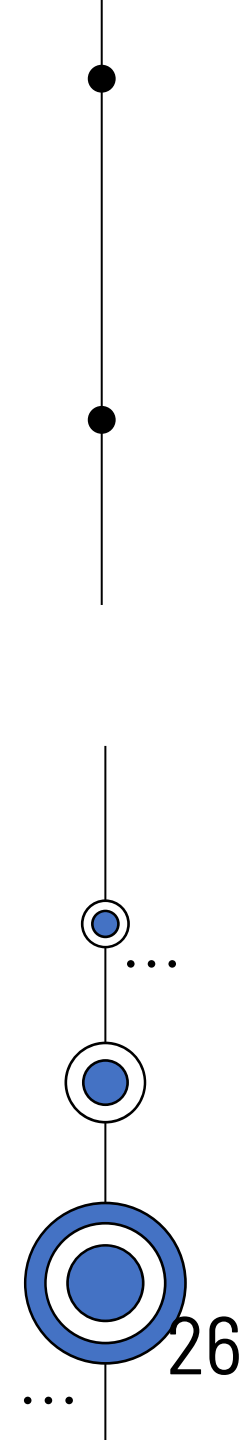
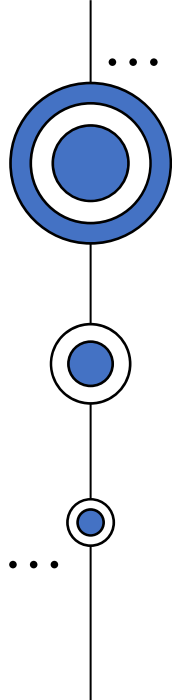
So Sánh



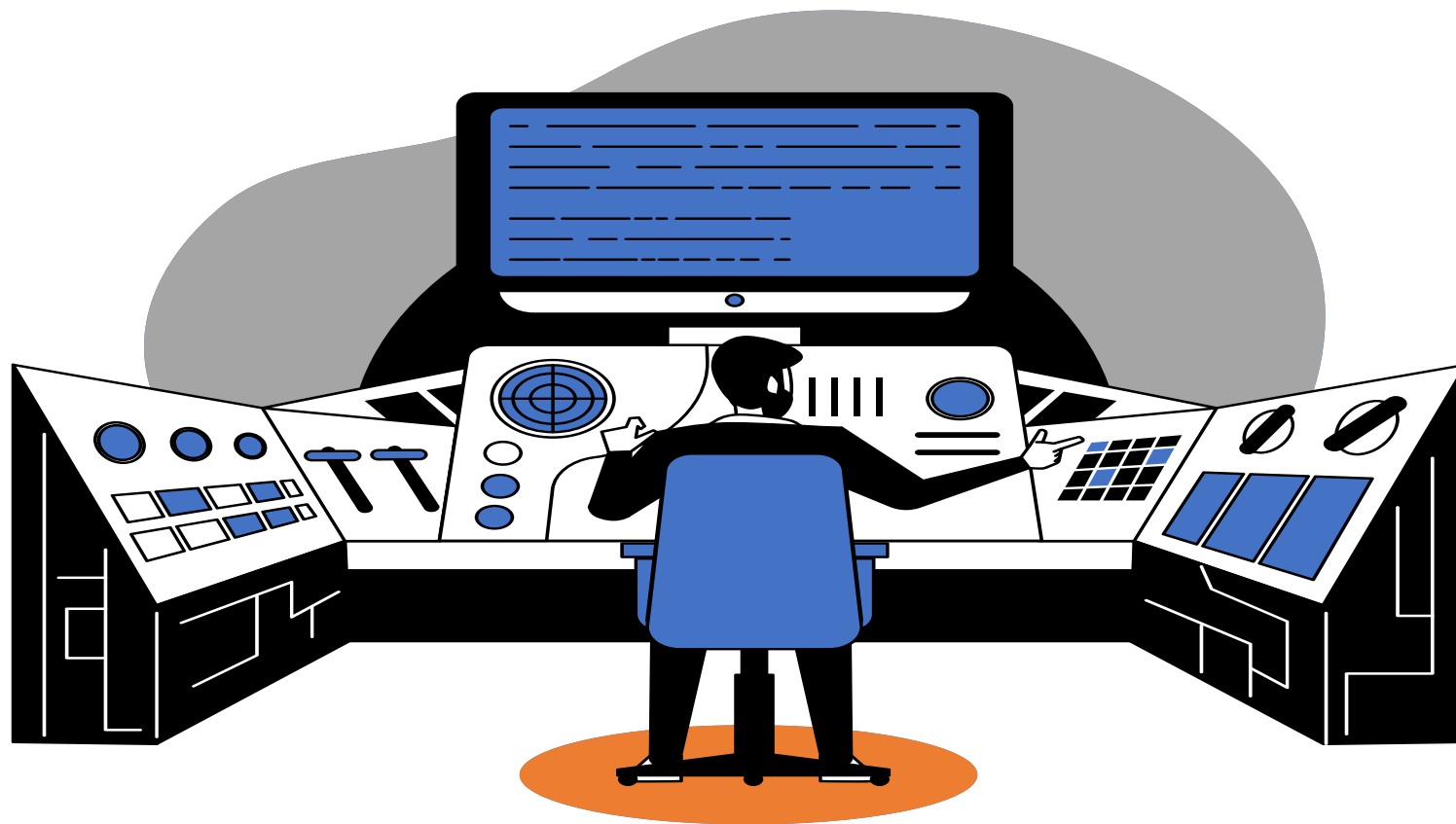
**5. IMAGE COMPRESSED BY
JPEG :QUALITY FACTOR =
15**

05

Demo code



DEMO NÉN ẢNH - PCA



Resources

- <https://machinelearningcoban.com/2017/06/15/pca/#1-gii-thiu>
- <https://machinelearningcoban.com/2017/06/21/pca2/>
- <https://miai.vn/2021/04/22/principal-component-analysis-pca-tuyet-c-hieu-giam-chieu-du-lieu/>
- <https://viblo.asia/p/cac-ky-thuat-dimensionality-reduction-OeVKB98A5kW>
- <https://daynhauhoc.com/t/pca-cho-tung-anh-trong-opencv/68351/15>
- <https://dataisg.org/tutorial/machine-learning/principal-component-analysis-pca/>

Thanks!

Do you have any questions?

19521270@gm.uit.edU.vn



CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), infographics & images by [Freepik](#) and illustrations by [Stories](#)

Please keep this slide for attribution