

# Bài tập PCA

## Bài tập 1

Cho ma trận  $A =$

```
[1, 2, 3],  
[4, 5, 6],  
[7, 8, 9],  
[2, 3, 1],  
[6, 3, 4],  
[5, 2, 7]
```

Với  $k = 2$  (số thành phần chính cần chọn)

Yêu cầu 1:

- Tính PCA bằng tay cho  $A$  theo  $k$
- In ra ma trận  $B$  với  $B = \text{PCA}(A)$

Yêu cầu 2:

- Tính PCA bằng thư viện (from sklearn.decomposition import PCA) cho  $A$  theo  $k$
- In ra ma trận  $B$  với  $B = \text{PCA}(A)$
- So sánh  $B$  của yêu cầu 1 và  $B$  của yêu cầu 2

## Bài tập 2

Giống Bài tập 1

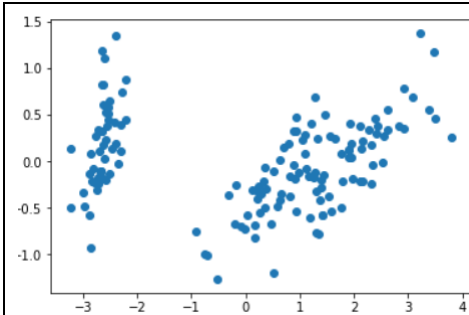
Nếu  $B$  của yêu cầu 1 và  $B$  của yêu cầu 2 **không giống nhau** thì xem tại sao và tìm cách giải quyết

## Bài tập 3

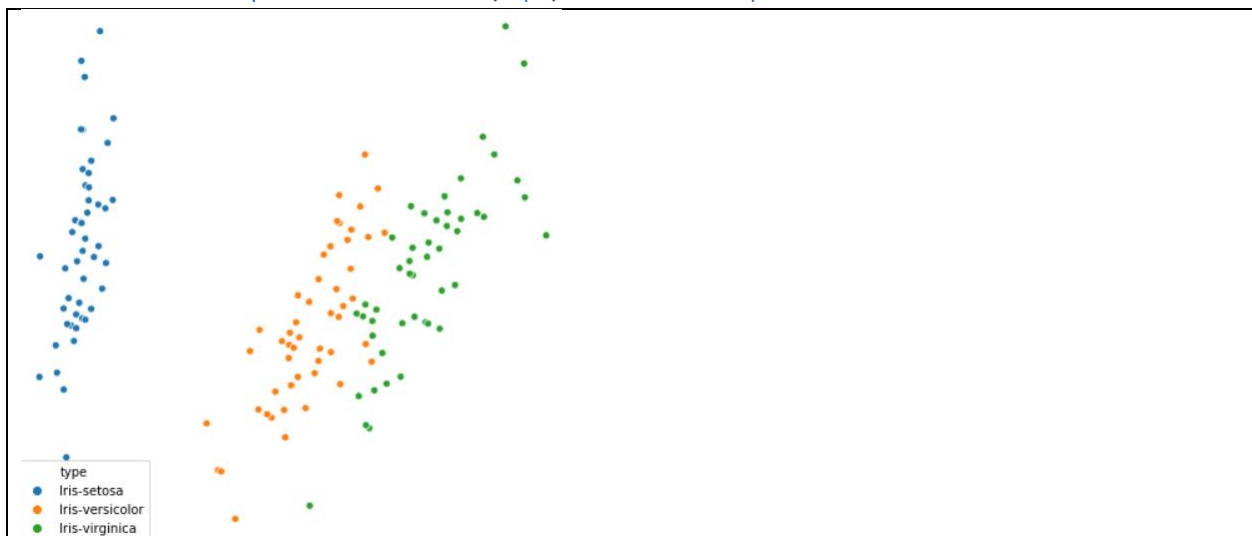
Đọc tập tin Iris.xls ra DataFrame data

Yêu cầu:

- Tính hệ số tương quan Pearson giữa các cột và cho nhận xét
- Tạo ma trận A gồm các cột 'sepalength', 'sepalwidth', 'petallength', 'petalwidth'
- Tính PCA cho A theo  $k=2 \rightarrow B = \text{PCA}(A)$
- Vẽ scatter plot B



- Vẽ scatter plot B có thêm loại (cột iris của data)



## Bài tập 4

Đọc tập tin hình ảnh original.png

Yêu cầu:

- Hiển thị hình
- Tính PCA theo  $k=100$
- Vẽ kết quả tính được bằng PCA