

# Báo cáo bài tập thực hành Xử lý đa chiều 02

Nguyễn Phú Thành - MSSV: 18110014

6/6/2021

**PROBLEM:** Apply Linear Discriminant Analysis algorithm on the Iris flower dataset. Make a comparison between LDA and PCA projected samples.

**Nhắc lại các bước của thuật toán LDA:**

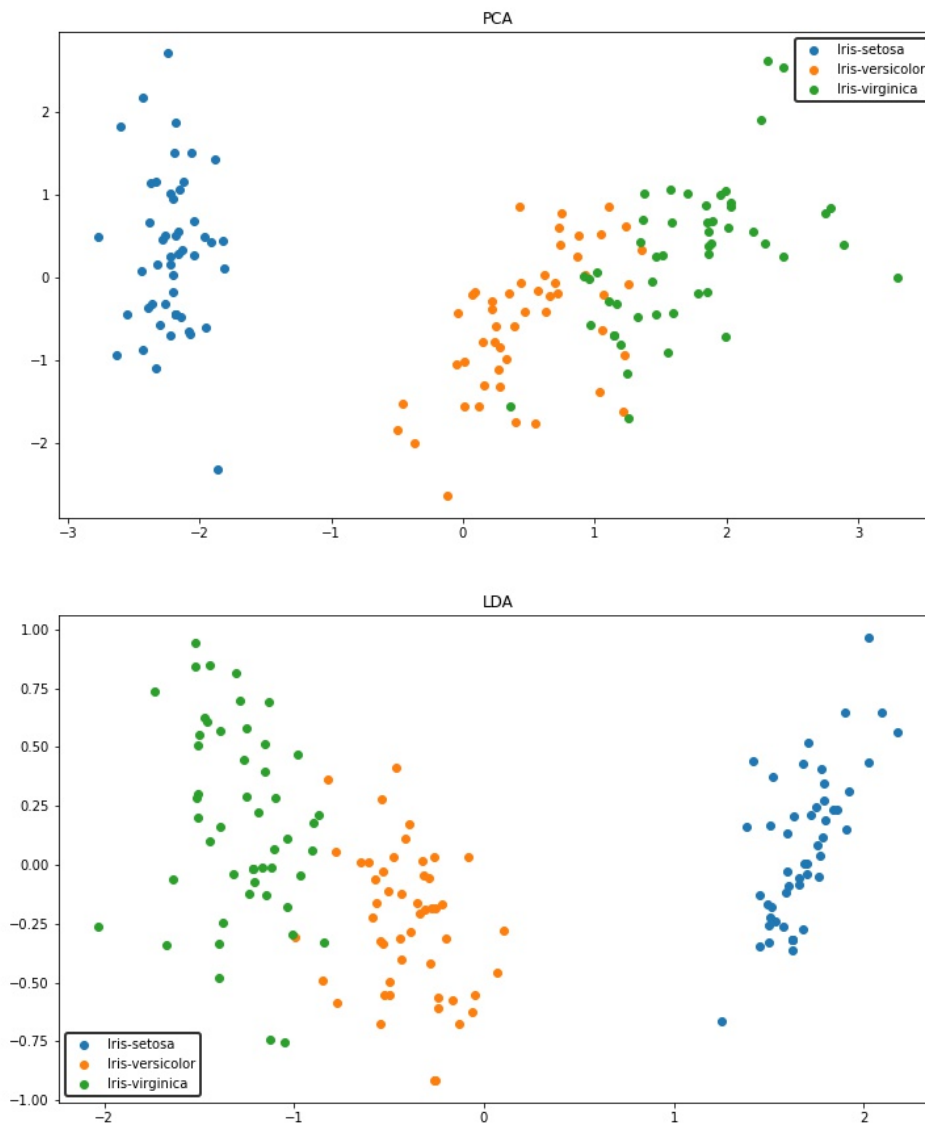
- Chuẩn hóa dữ liệu  $\mathbf{X}$  ban đầu với  $d$  chiều, tức thực hiện biến đổi:

$$\frac{x_{nj} - \bar{x}_j}{\sqrt{s_{jj}}}, \forall j = 1, 2, \dots, d$$

trong đó  $x_{nj}$  là thành phần thứ  $j$  của quan sát thứ  $n$ ,  $\bar{x}_j, s_{jj}$  lần lượt là phần tử thứ  $j$  của vectơ trung bình mẫu  $d$  chiều và phần tử thứ  $j$  nằm trên đường chéo chính của ma trận hiệp phương sai mẫu

- Với mỗi lớp, tính vectơ trung bình  $\boldsymbol{\mu}_i = \frac{1}{N_i} \sum_{\mathbf{x} \in \text{Lớp thứ } i} \mathbf{x}, \forall i = 1, 2, \dots, C$
- Tính within-class scatter matrix:  $\mathbf{S}_W = \sum_{i=1}^C \mathbf{S}_i$  trong đó  $\mathbf{S}_i = \sum_{\mathbf{x} \in \text{Lớp thứ } i} (\mathbf{x} - \boldsymbol{\mu}_i)(\mathbf{x} - \boldsymbol{\mu}_i)^T$
- Tính between-class scatter matrix:  $\mathbf{S}_B = \sum_{i=1}^C N_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T$
- Tính các trị riêng, vectơ riêng của ma trận  $\mathbf{S}_W^{-1} \mathbf{S}_B$
- Sắp xếp các trị riêng theo chiều giảm dần
- Chọn ra  $k < d$  vectơ riêng ứng với  $k$  trị riêng đầu tiên, xây dựng ma trận chiếu  $\mathbf{W} \in \mathbb{R}^{d \times k}$  trong đó mỗi cột là một vectơ riêng
- Dữ liệu sau khi chiếu xuống không gian  $k$  chiều bởi ma trận  $\mathbf{W}$  là:  $\mathbf{XW}$

**Kết quả thuật toán LDA trên tập dữ liệu Iris và so sánh với kết quả của PCA**



Hình 1: So sánh LDA và PCA trên bộ dữ liệu Iris

#### Nhận xét:

- **Điểm chung:** Lớp hoa iris-setosa bởi 2 thuật toán LDA và PCA đều phân tách hoàn toàn so với hai lớp hoa còn lại
- **Điểm khác:**

1. Kết quả giữa PCA và LDA có vẻ đối xứng với nhau qua trục tung
2. Kết quả của LDA phân tách hai lớp Iris-versicolor và Iris-virginica tốt hơn so với kết quả của PCA