ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/348885432

Fisher and Linear Discriminant Analysis

Presentation · January 2021

CITATION
1

READS
216

2 authors:

Benyamin Ghojogh
University of Waterloo
91 PUBLICATIONS **158** CITATIONS

SEE PROFILE

Mark Crowley
University of Waterloo
111 PUBLICATIONS **314** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Physics and Machine Learning View project

Project Metaheuristic Optimization Algorithms View project

All content following this page was uploaded by Benyamin Ghojogh on 29 January 2021.

The user has requested enhancement of the downloaded file.

Fisher and Linear Discriminant Analysis

Department of Electrical & Computer Engineering,
University of Waterloo, ON, Canada

Data and Knowledge Modeling and Analysis (ECE 657A)
Course Instructor: Prof. Mark Crowley
TA and Presenter of Slides: Benyamin Ghojogh

Dataset Notations

$$\text{training dataset: } \{\boldsymbol{x}_i \in \mathbb{R}^d\}_{i=1}^n, \boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n] \in \mathbb{R}^{d \times n} \tag{1}$$

$$j\text{-th training class: } \{\boldsymbol{x}_i^{(j)} \in \mathbb{R}^d\}_{i=1}^{n_j}, \quad \forall j \in \{1, \dots, c\} \tag{2}$$

$$\text{mean of training data: } \mathbb{R}^d \ni \boldsymbol{\mu}_x := \frac{1}{n} \sum_{i=1}^n \boldsymbol{x}_i \tag{3}$$

$$\text{test dataset: } \{\boldsymbol{x}_{t,i} \in \mathbb{R}^d\}_{i=1}^{n_t}, \boldsymbol{X}_t = [\boldsymbol{x}_{t,1}, \dots, \boldsymbol{x}_{t,n_t}] \in \mathbb{R}^{d \times n_t} \tag{4}$$

Lecture Outline

FDA: Inter- and Intra-Class Scatters

$$\mathbb{R}^d \ni \boldsymbol{\mu}_j := \frac{1}{n_j} \sum_{i=1}^{n_j} \boldsymbol{x}_i^{(j)} \tag{5}$$

$$\mathbb{R}^d \ni \boldsymbol{\mu} := \frac{1}{\sum_{k=1}^{c} n_k} \sum_{j=1}^{c} n_j \, \boldsymbol{\mu}_j = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \tag{6}$$

$$\mathbb{R}^{d \times d} \ni \boldsymbol{S}_B := \sum_{j=1}^{c} n_j (\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^\top \tag{7}$$

$$\mathbb{R}^{d \times d} \ni \boldsymbol{S}_W := \sum_{j=1}^{c} \boldsymbol{S}_j = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (\boldsymbol{x}_i^{(j)} - \boldsymbol{\mu}_j)(\boldsymbol{x}_i^{(j)} - \boldsymbol{\mu}_j)^\top \tag{8}$$

Lecture Outline

Scatters of Projections

$$\widehat{\boldsymbol{x}} = \boldsymbol{u}\boldsymbol{u}^\top \breve{\boldsymbol{x}} \tag{9}$$

$$\begin{aligned} ||\widehat{\boldsymbol{x}}||_2^2 &= ||\boldsymbol{u}\boldsymbol{u}^\top \breve{\boldsymbol{x}}||_2^2 = (\boldsymbol{u}\boldsymbol{u}^\top \breve{\boldsymbol{x}})^\top (\boldsymbol{u}\boldsymbol{u}^\top \breve{\boldsymbol{x}}) \\ &= \breve{\boldsymbol{x}}^\top \boldsymbol{u} \underbrace{\boldsymbol{u}^\top \boldsymbol{u}}_{1} \boldsymbol{u}^\top \breve{\boldsymbol{x}} = \breve{\boldsymbol{x}}^\top \boldsymbol{u}\, \boldsymbol{u}^\top \breve{\boldsymbol{x}} = \boldsymbol{u}^\top \breve{\boldsymbol{x}}\, \breve{\boldsymbol{x}}^\top \boldsymbol{u} \end{aligned} \tag{10}$$

$$\sum_{i=1}^{n} ||\widehat{\boldsymbol{x}}_i||_2^2 = \sum_{i=1}^{n} \boldsymbol{u}^\top \breve{\boldsymbol{x}}_i\, \breve{\boldsymbol{x}}_i^\top \boldsymbol{u} = \boldsymbol{u}^\top \Big(\sum_{i=1}^{n} \breve{\boldsymbol{x}}_i\, \breve{\boldsymbol{x}}_i^\top \Big) \boldsymbol{u} \tag{11}$$

$$\text{covariance matrix: } \mathbb{R}^{d\times d} \ni \boldsymbol{S} := \sum_{i=1}^{n} \breve{\boldsymbol{x}}_i\, \breve{\boldsymbol{x}}_i^\top = \breve{\boldsymbol{X}} \breve{\boldsymbol{X}}^\top$$

$$||\widehat{\boldsymbol{X}}||_F^2 = \boldsymbol{u}^\top \boldsymbol{S} \boldsymbol{u}, \quad ||\widehat{\boldsymbol{X}}||_F^2 = \mathbf{tr}(\boldsymbol{U}^\top \boldsymbol{S}\, \boldsymbol{U}). \tag{12}$$

Projection & Reconstruction

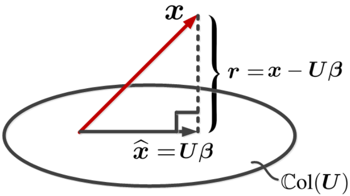
$$\text{projection: } \mathbb{R}^p \ni \widetilde{\boldsymbol{x}} := \boldsymbol{U}^\top \boldsymbol{x}, \widetilde{\boldsymbol{X}} = [\widetilde{\boldsymbol{x}}_1, \dots, \widetilde{\boldsymbol{x}}_n] \in \mathbb{R}^{p \times n} \tag{13}$$

$$\text{reconstruction: } \mathbb{R}^d \ni \widehat{\boldsymbol{x}} := \boldsymbol{U}\boldsymbol{U}^\top \boldsymbol{x} + \boldsymbol{\mu}_x = \boldsymbol{U}\widetilde{\boldsymbol{x}} + \boldsymbol{\mu}_x \tag{14}$$

$$\widehat{\boldsymbol{X}} = [\widehat{\boldsymbol{x}}_1, \dots, \widehat{\boldsymbol{x}}_n] \in \mathbb{R}^{d \times n}$$

$$\text{test projection: } \mathbb{R}^{p \times n_t} \ni \widetilde{\boldsymbol{X}}_t = \boldsymbol{U}^\top \boldsymbol{X}_t, \tag{15}$$

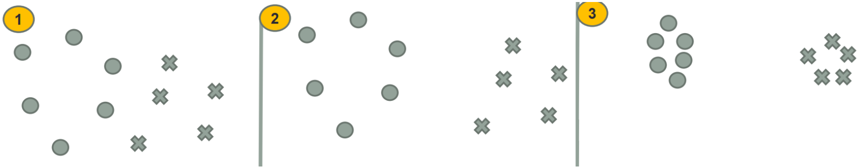
$$\text{test reconst.: } \mathbb{R}^{d \times n_t} \ni \widehat{\boldsymbol{X}}_t = \boldsymbol{U}\boldsymbol{U}^\top \boldsymbol{X}_t + \boldsymbol{\mu}_x = \boldsymbol{U}\widetilde{\boldsymbol{X}}_t + \boldsymbol{\mu}_x, \tag{16}$$



FDA Goal

For better **separation** of classes in some data (Fig 1):

- Make **inter-class scatter** larger (Fig 2)
- Make **intra-class scatter** smaller (Fig 3)



FDA does this for **supervised** subspace learning.
The first FDA paper: [1] by Ronald A. Fisher (1890 – 1962)

FDA Optimization

$$\underset{\boldsymbol{U}}{\text{maximize}} \quad f(\boldsymbol{U}) := \frac{\textbf{tr}(\boldsymbol{U}^\top \boldsymbol{S}_B\, \boldsymbol{U})}{\textbf{tr}(\boldsymbol{U}^\top \boldsymbol{S}_W\, \boldsymbol{U})}. \tag{17}$$

$$\begin{aligned} &\underset{\boldsymbol{U}}{\text{maximize}} && \textbf{tr}(\boldsymbol{U}^\top \boldsymbol{S}_B\, \boldsymbol{U}) \\ &\text{subject to} && \boldsymbol{U}^\top \boldsymbol{S}_W\, \boldsymbol{U} = \boldsymbol{I}. \end{aligned} \tag{18}$$

$$\mathcal{L} = \textbf{tr}(\boldsymbol{U}^\top \boldsymbol{S}_B\, \boldsymbol{U}) - \textbf{tr}\big(\Lambda^\top (\boldsymbol{U}^\top \boldsymbol{S}_W\, \boldsymbol{U} - \boldsymbol{I})\big) \tag{19}$$

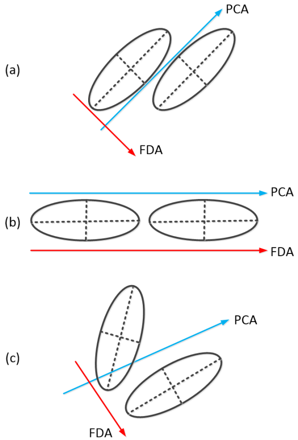
$$\begin{aligned} &\mathbb{R}^{d\times p} \ni \frac{\partial \mathcal{L}}{\partial \boldsymbol{U}} = 2\,\boldsymbol{S}_B\, \boldsymbol{U} - 2\,\boldsymbol{S}_W\, \boldsymbol{U}\Lambda \overset{\text{set}}{=} 0 \\ &\implies 2\,\boldsymbol{S}_B\, \boldsymbol{U} = 2\,\boldsymbol{S}_W\, \boldsymbol{U}\Lambda \implies \boxed{\boldsymbol{S}_B\, \boldsymbol{U} = \boldsymbol{S}_W\, \boldsymbol{U}\Lambda} \end{aligned} \tag{20}$$

FDA: Rank and Dimensionality of Subspace

$$\boldsymbol{S}_B\, \boldsymbol{U} = \boldsymbol{S}_W\, \boldsymbol{U}\Lambda \implies \boxed{\boldsymbol{S}_W^{-1}\boldsymbol{S}_B\, \boldsymbol{U} = \boldsymbol{U}\Lambda} \tag{21}$$

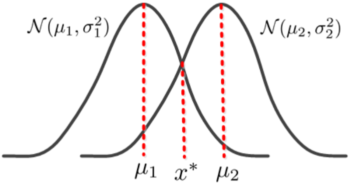
$$\begin{aligned}
&\textbf{rank}(\boldsymbol{S}_W^{-1}\boldsymbol{S}_B) \leq \min\big(\textbf{rank}(\boldsymbol{S}_W^{-1}), \textbf{rank}(\boldsymbol{S}_B)\big) \\
&\leq \min\big(\min(d, n-1), \min(d, c-1)\big) \\
&= \min(d, n-1, c-1) = \boxed{c-1}
\end{aligned} \tag{22}$$

FDA vs PCA



Lecture Outline

LDA: Optimization for Boundary of Classes



$$\begin{aligned}\mathbb{P}(\text{error}) = \; & \mathbb{P}(x > x^* \,|\, x \in \mathcal{C}_1)\, \mathbb{P}(x \in \mathcal{C}_1) \\ & + \mathbb{P}(x < x^* \,|\, x \in \mathcal{C}_2)\, \mathbb{P}(x \in \mathcal{C}_2)\end{aligned}$$

$$\mathbb{P}(x < c, x \in \mathcal{C}_1) = F_1(c) \implies \mathbb{P}(x > x^*, x \in \mathcal{C}_1) = 1 - F_1(x^*)$$

$$\mathbb{P}(x < x^*, x \in \mathcal{C}_2) = F_2(x^*)$$

$$\mathbb{P}(x \in \mathcal{C}_1) = f_1(x) = \pi_1, \mathbb{P}(x \in \mathcal{C}_2) = f_2(x) = \pi_2$$

$$\underset{x^*}{\text{minimize}} \quad \mathbb{P}(\text{error}) \implies \underset{x^*}{\text{minimize}} \quad \big(1 - F_1(x^*)\big)\, \pi_1 + F_2(x^*)\, \pi_2$$

LDA: Optimization for Boundary of Classes

$$\underset{x^*}{\text{minimize}} \quad \mathbb{P}(\text{error}) \implies \underset{x^*}{\text{minimize}} \quad \big(1 - F_1(x^*)\big)\,\pi_1 + F_2(x^*)\,\pi_2$$

$$\frac{\partial\,\mathbb{P}(\text{error})}{\partial x^*} = -f_1(x^*)\,\pi_1 + f_2(x^*)\,\pi_2 \overset{\text{set}}{=} 0$$

$$\implies \boxed{f_1(x^*)\,\pi_1 = f_2(x^*)\,\pi_2} \tag{23}$$

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\Big(-\,\frac{(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2}\Big) \tag{24}$$

$$\frac{1}{\sqrt{(2\pi)^d|\Sigma_1|}} \exp\Big(-\,\frac{(\boldsymbol{x}-\boldsymbol{\mu}_1)^\top \Sigma_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)}{2}\Big)\,\pi_1$$

$$= \frac{1}{\sqrt{(2\pi)^d|\Sigma_2|}} \exp\Big(-\,\frac{(\boldsymbol{x}-\boldsymbol{\mu}_2)^\top \Sigma_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)}{2}\Big)\,\pi_2,$$

Lecture Outline

LDA: Binary Classes

$$\Sigma_1 = \Sigma_2 = \Sigma \tag{25}$$

$$\frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\Big(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_1)^\top \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_1)}{2}\Big)\, \pi_1$$

$$= \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\Big(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_2)^\top \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_2)}{2}\Big)\, \pi_2, \tag{26}$$

$$\implies \exp\Big(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_1)^\top \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_1)}{2}\Big)\, \pi_1$$

$$= \exp\Big(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_2)^\top \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_2)}{2}\Big)\, \pi_2 \tag{27}$$

LDA: Binary Classes

$$\exp\Big(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_1)^\top\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)}{2}\Big)\,\pi_1$$

$$=\exp\Big(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_2)^\top\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)}{2}\Big)\,\pi_2, \tag{28}$$

$$\implies -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^\top\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)+\ln(\pi_1) \tag{29}$$

$$=-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_2)^\top\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)+\ln(\pi_2), \tag{30}$$

$$(\boldsymbol{x}-\boldsymbol{\mu}_1)^\top\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)=(\boldsymbol{x}^\top-\boldsymbol{\mu}_1^\top)\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)$$

$$=\boldsymbol{x}^\top\Sigma^{-1}\boldsymbol{x}-\boldsymbol{x}^\top\Sigma^{-1}\boldsymbol{\mu}_1-\boldsymbol{\mu}_1^\top\Sigma^{-1}\boldsymbol{x}+\boldsymbol{\mu}_1^\top\Sigma^{-1}\boldsymbol{\mu}_1$$

$$=\boldsymbol{x}^\top\Sigma^{-1}\boldsymbol{x}+\boldsymbol{\mu}_1^\top\Sigma^{-1}\boldsymbol{\mu}_1-2\,\boldsymbol{\mu}_1^\top\Sigma^{-1}\boldsymbol{x}, \tag{31}$$

LDA: Binary Classes

$$
\begin{aligned}
(\boldsymbol{x}-\boldsymbol{\mu}_1)^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) &= \boldsymbol{x}^\top \Sigma^{-1}\boldsymbol{x} + \boldsymbol{\mu}_1^\top \Sigma^{-1}\boldsymbol{\mu}_1 - 2\,\boldsymbol{\mu}_1^\top \Sigma^{-1}\boldsymbol{x}, \\
\implies -\frac{1}{2}\boldsymbol{x}^\top \Sigma^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_1^\top \Sigma^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_1^\top \Sigma^{-1}\boldsymbol{x} + \ln(\pi_1) & \\
= -\frac{1}{2}\boldsymbol{x}^\top \Sigma^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_2^\top \Sigma^{-1}\boldsymbol{\mu}_2 + \boldsymbol{\mu}_2^\top \Sigma^{-1}\boldsymbol{x} + \ln(\pi_2) & \qquad (32)
\end{aligned}
$$

$\implies$

$$
\boxed{\begin{aligned} & 2\left(\Sigma^{-1}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)\right)^\top \boldsymbol{x} \\ & \quad + \left(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2\right)^\top \Sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\big) + 2\,\ln(\frac{\pi_2}{\pi_1}) = 0 \end{aligned}} \qquad (33)
$$

LDA: Binary Classes

$$
\begin{aligned}
\delta(\boldsymbol{x}) := 2 \left(\Sigma^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)\right)^\top \boldsymbol{x} \\
+ \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right) + 2 \ln(\frac{\pi_2}{\pi_1})
\end{aligned} \tag{34}
$$

$$
\widehat{\mathcal{C}}(x) = \begin{cases} 1, & \text{if } \delta(\boldsymbol{x}) < 0, \\ 2, & \text{if } \delta(\boldsymbol{x}) > 0. \end{cases} \tag{35}
$$

LDA: Multiple Classes

$$f_k(\boldsymbol{x})\,\pi_k = \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} \exp\Big(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_k)^\top \Sigma_k^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_k)}{2}\Big)\pi_k. \tag{36}$$

$$\begin{aligned}\ln(f_k(\boldsymbol{x})\,\pi_k) = &-\frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Sigma_k|) \\ &- \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^\top \Sigma_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k) + \ln(\pi_k)\end{aligned} \tag{37}$$

$$\delta_k(\boldsymbol{x}) := -\frac{1}{2}\ln(|\Sigma_k|) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^\top \Sigma_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k) + \ln(\pi_k) \tag{38}$$

$$\widehat{\mathcal{C}}(\boldsymbol{x}) = \arg\max_k\ \delta_k(\boldsymbol{x}) \tag{39}$$

Lecture Outline

Relation of FDA and LDA

$$d^2_{\boldsymbol{A}}(\boldsymbol{x}, \boldsymbol{\mu}_k) := ||\boldsymbol{x} - \boldsymbol{\mu}_k||^2_{\boldsymbol{A}} = (\boldsymbol{x} - \boldsymbol{\mu}_k)^\top \boldsymbol{A} (\boldsymbol{x} - \boldsymbol{\mu}_k) \tag{40}$$

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{U}^\top \succeq 0 \tag{41}$$

$$\begin{aligned} ||\boldsymbol{x} - \boldsymbol{\mu}_k||^2_{\boldsymbol{A}} &= (\boldsymbol{x} - \boldsymbol{\mu}_k)^\top \boldsymbol{U}\boldsymbol{U}^\top (\boldsymbol{x} - \boldsymbol{\mu}_k) \\ &= (\boldsymbol{U}^\top \boldsymbol{x} - \boldsymbol{U}^\top \boldsymbol{\mu}_k)^\top (\boldsymbol{U}^\top \boldsymbol{x} - \boldsymbol{U}^\top \boldsymbol{\mu}_k) \end{aligned} \tag{42}$$

$$\boldsymbol{x} \mapsto \boldsymbol{u}^\top \boldsymbol{x}, \quad \boldsymbol{\mu} \mapsto \boldsymbol{u}^\top \boldsymbol{\mu}, \quad \Sigma \mapsto \boldsymbol{u}^\top \Sigma\, \boldsymbol{u} \tag{43}$$

$$f := \frac{\sigma_b^2}{\sigma_w^2} = \frac{(\boldsymbol{u}^\top \boldsymbol{\mu}_2 - \boldsymbol{u}^\top \boldsymbol{\mu}_1)^2}{\boldsymbol{u}^\top \Sigma_2\, \boldsymbol{u} + \boldsymbol{u}^\top \Sigma_1\, \boldsymbol{u}} = \frac{\big(\boldsymbol{u}^\top (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)\big)^2}{\boldsymbol{u}^\top (\Sigma_2 + \Sigma_1)\, \boldsymbol{u}}. \tag{44}$$

Relation of FDA and LDA

$$\underset{\boldsymbol{u}}{\text{maximize}} \quad \frac{\big(\boldsymbol{u}^\top(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)\big)^2}{\boldsymbol{u}^\top(\Sigma_2 + \Sigma_1)\,\boldsymbol{u}}, \tag{45}$$

$$\begin{aligned} &\underset{\boldsymbol{u}}{\text{maximize}} && \big(\boldsymbol{u}^\top(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)\big)^2, \\ &\text{subject to} && \boldsymbol{u}^\top(\Sigma_2 + \Sigma_1)\,\boldsymbol{u} = 1, \end{aligned} \tag{46}$$

$$\mathcal{L} = \big(\boldsymbol{u}^\top(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)\big)^2 - \lambda\big(\boldsymbol{u}^\top(\Sigma_2 + \Sigma_1)\,\boldsymbol{u} - 1\big) \tag{47}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{u}} = 2\,(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^\top \boldsymbol{u} - 2\,\lambda\,(\Sigma_2 + \Sigma_1)\,\boldsymbol{u} \overset{\text{set}}{=} 0$$

$$\implies (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^\top\,\boldsymbol{u} = \lambda\,(\Sigma_2 + \Sigma_1)\,\boldsymbol{u}, \tag{48}$$

$$\boldsymbol{u} \propto (\Sigma_2 + \Sigma_1)^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^\top \tag{49}$$

Relation of FDA and LDA

$$\boldsymbol{u} \propto (\Sigma_2 + \Sigma_1)^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^\top \tag{50}$$

$$\text{In LDA: } \Sigma_1 = \Sigma_2 \tag{51}$$

$$\boldsymbol{u} \propto (2\,\Sigma)^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^\top \propto \Sigma^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^\top \tag{52}$$

$$\boldsymbol{u}^\top \boldsymbol{x} \propto \big(\Sigma^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^\top\big)^\top \boldsymbol{x} \tag{53}$$

Compare with Eq.(34) $\implies$ Upto a scaling factor:

$$\boxed{\text{LDA} \equiv \text{FDA}} \tag{54}$$

Note that FDA is subspace learning but LDA is classifier (but can be seen as metric learning).

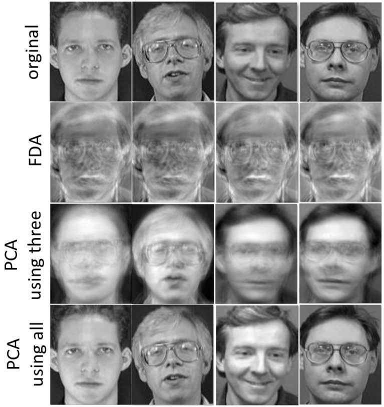
Lecture Outline

Fisherfaces

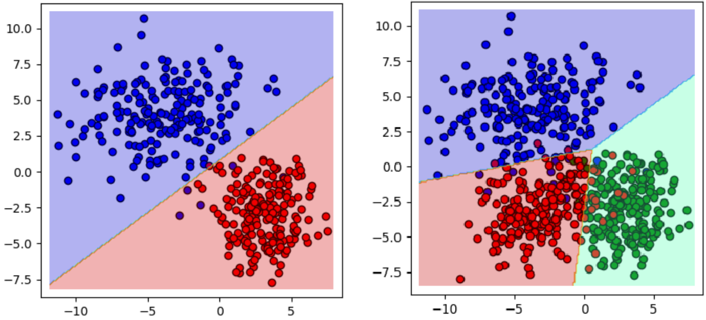
Fisherfaces vs. Eigenfaces on ORL face dataset:

Reconstruction: FDA vs. PCA

Reconstructions by FDA and PCA:

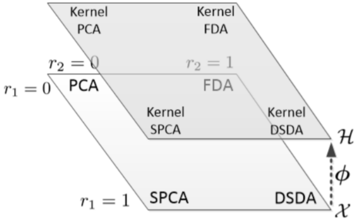


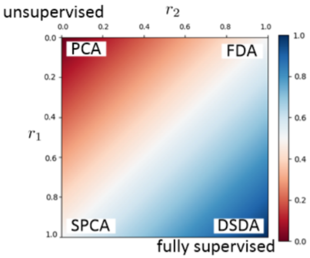
LDA Examples



Additional Note: Roweis Discriminant Analysis

- A **generalized** subspace learning method based on **generalized eigenvalue problem**.
- The **Roweis map** generalizes PCA, supervised PCA, and FDA:
- Paper: [2, 3]





Useful Resources To Read

- Tutorial paper: "Linear and Quadratic Discriminant Analysis: Tutorial" [4]
- Tutorial paper: "Fisher and Kernel Fisher Discriminant Analysis: Tutorial" [5]
- Tutorial paper: "Eigenvalue and generalized eigenvalue problems: Tutorial" [6]
- Tutorial YouTube videos by Prof. Ali Ghodsi at University of Waterloo: [Click here] and [Click here]

References

[1] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[2] B. Ghojogh, F. Karray, and M. Crowley, "Roweis discriminant analysis: A generalized subspace learning method," *arXiv preprint arXiv:1910.05437*, 2019.

[3] B. Ghojogh, F. Karray, and M. Crowley, "Generalized subspace learning by Roweis discriminant analysis," in *International Conference on Image Analysis and Recognition*, pp. 328–342, Springer, 2020.

[4] B. Ghojogh and M. Crowley, "Linear and quadratic discriminant analysis: Tutorial," *arXiv preprint arXiv:1906.02590*, 2019.

[5] B. Ghojogh, F. Karray, and M. Crowley, "Fisher and kernel Fisher discriminant analysis: Tutorial," *arXiv preprint arXiv:1906.09436*, 2019.

[6] B. Ghojogh, F. Karray, and M. Crowley, "Eigenvalue and generalized eigenvalue problems: Tutorial," *arXiv preprint arXiv:1903.11240*, 2019.

View publication stats