Análise de Discriminante Linear

Linear Discriminant Analysis - LDA

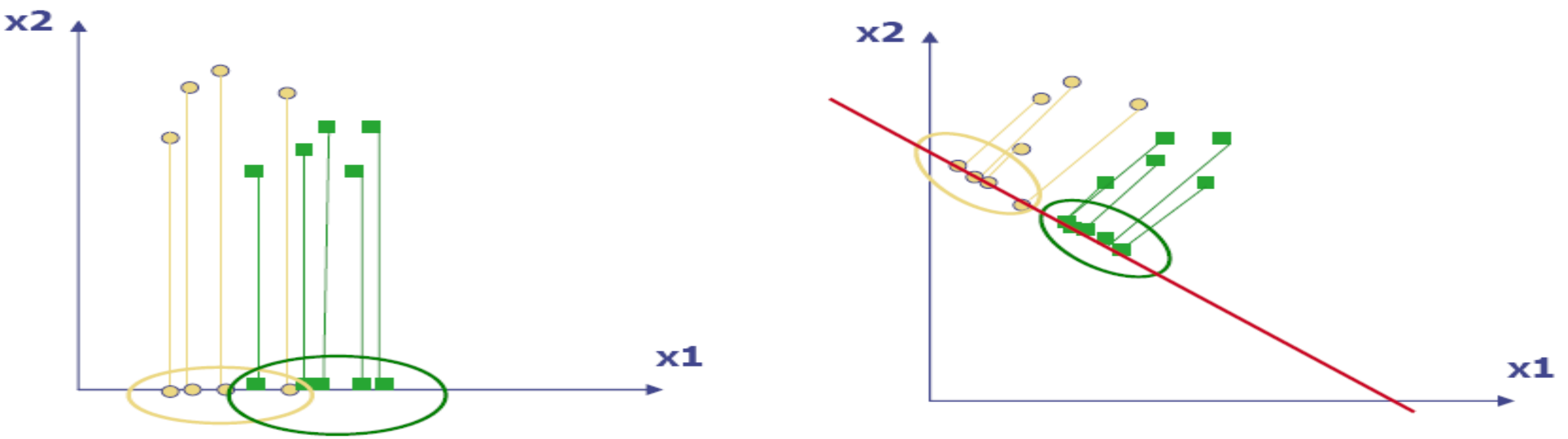
Elias R. da Silva Jr
Tsang Ing Ren
George Darmiton da Cunha Cavalcanti
CIn/UFPE

Roteiro

- Introdução
- LDA
- PCA x LDA
- Implementação

Introdução

- Técnicas de extração de características
  - Operam sobre o espaço de características de alta dimensionalidade
  - projetando-o sobre um espaço de características de menor dimensionalidade
  - De modo a salientar relações que:
    - representem o espaço anterior
    - sem causar prejuízo a discriminação entre os padrões

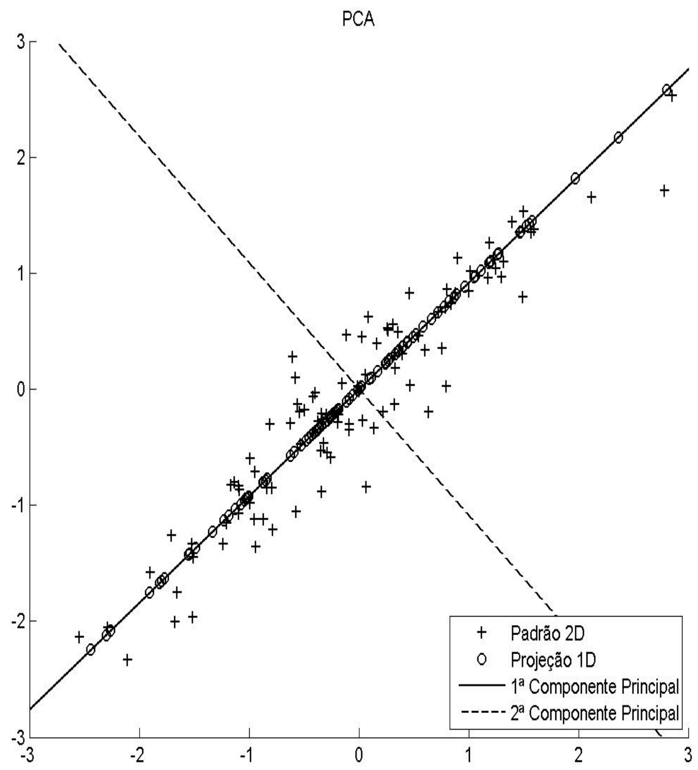


Poor Projection   Good Projection

Introdução

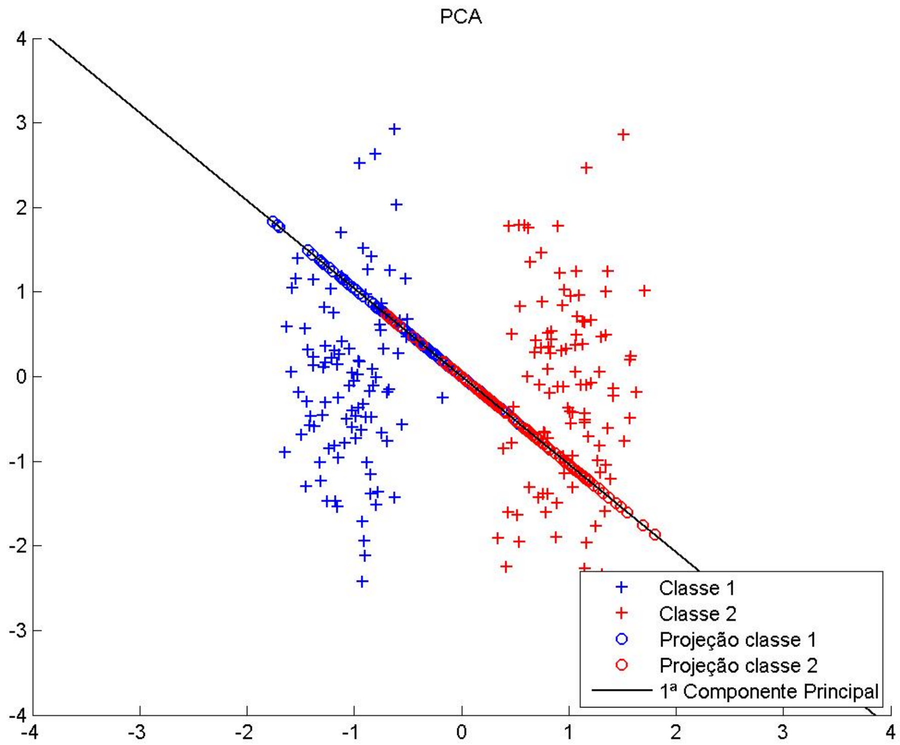
PCA

- Importante técnica de extração de características
- Não-supervisionada
- Visa encontrar uma redução de dimensionalidade projetada sobre as maiores variâncias dos dados



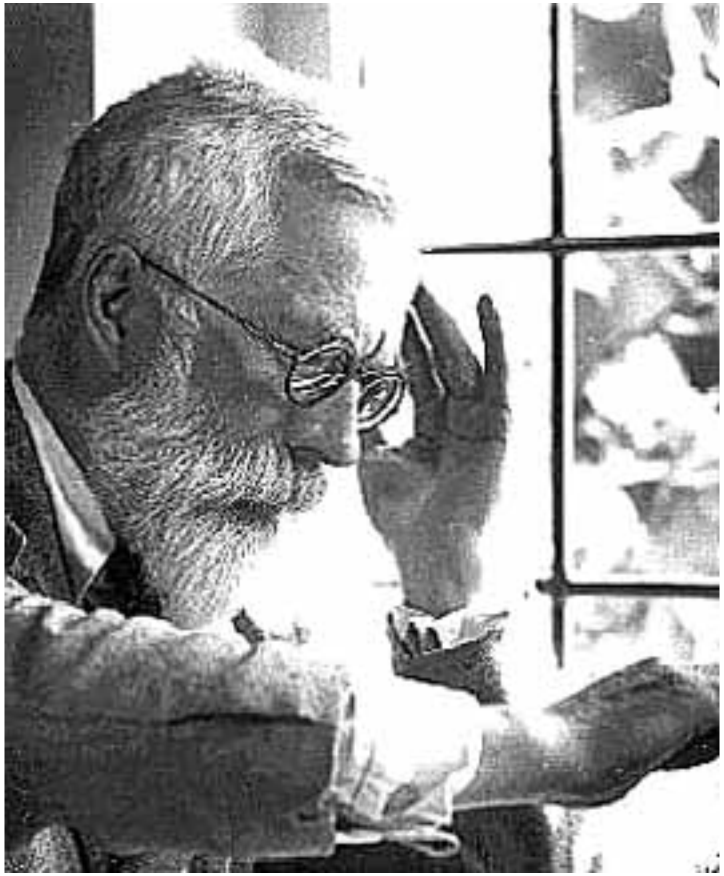
Introdução

- Mas o que fazer quando o problema alvo é supervisionado e o PCA não responde bem a ele?



LDA

- Linear Discriminant Analysis foi proposta por Ronald A. Fisher em 1936 [1]
- Conhecida por:
  - Linear Discriminant Analysis (LDA)
  - Fisher Discriminant Analysis (FDA)
  - Fisher’s Linear Discriminant (FLD)
  - Fisher’s Linear Discriminant Analysis (FLDA)



LDA

O que é LDA?

- É uma técnica de extração de características
- Supervisionada
- Amplamente utilizada e bem sucedida em vários problemas de reconhecimento de padrões.
  - Reconhecimento de face [2]
  - Reconhecimento de emoções [3]
  - Verificação on-line de assinaturas [4]

LDA

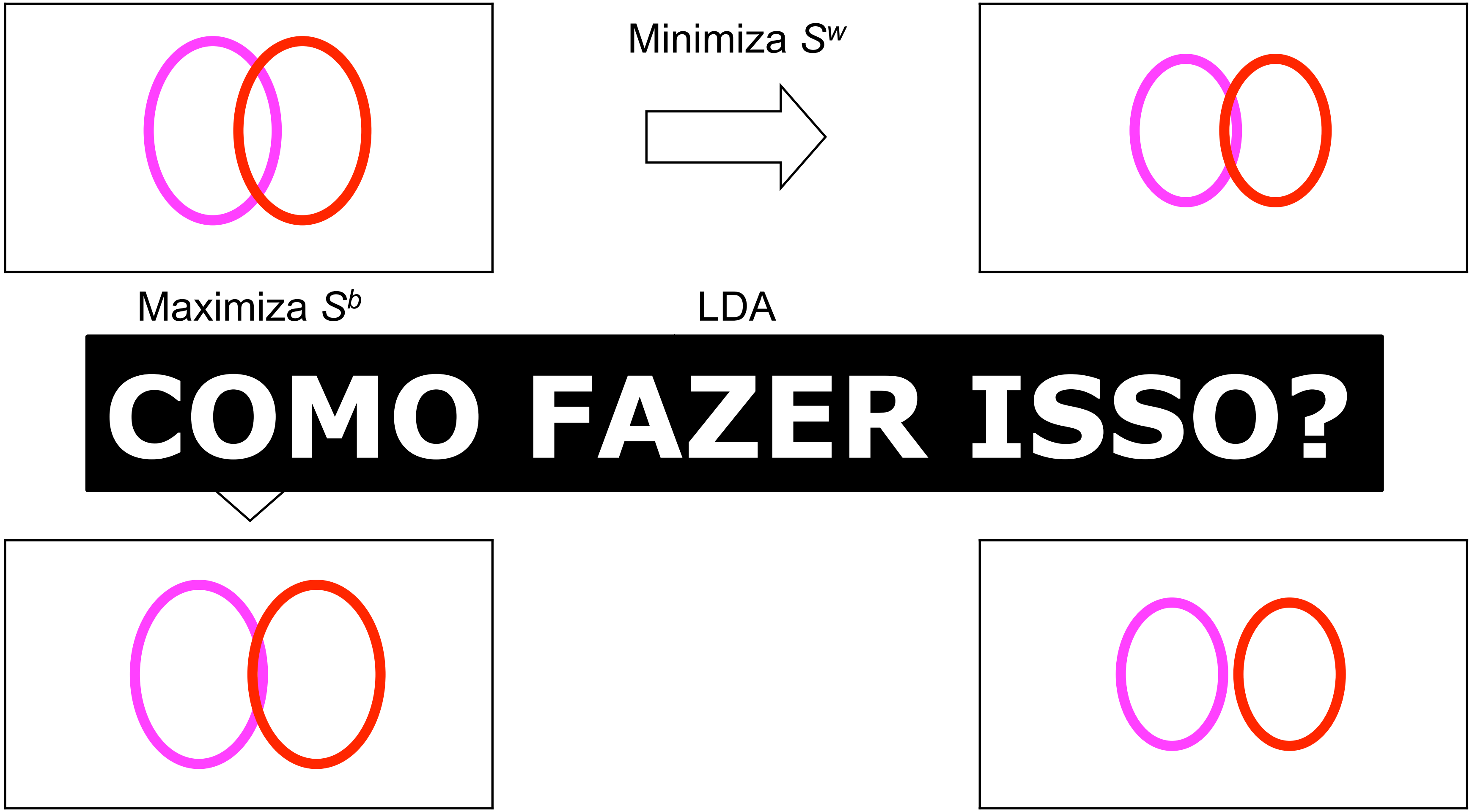
Objetivos gerais

- Desempenhar redução de dimensionalidade
  - Simplificação do problema
  - Redução do custo computacional
- Manter (ou melhorar) a capacidade de discriminação entre as classes do problemas

Objetivos específicos

- Minimizar a dispersão intraclasse
  - Padrões de uma mesma classe estejam agrupados o mais densamente possível
- Maximizar a dispersão entreclasses
  - Agrupamentos de classes distintas sejam o mais separados possível

LDA



COMO FAZER ISSO?

LDA

- Deste modo, e diferente do PCA, o LDA utiliza duas matrizes de dispersão distintas na implementação de seus objetivos:
  - A matriz de dispersão intraclasse (S^w)
    - Computa a variância dos padrões em relação a classe a qual pertence
  - A matriz de dispersão entreclasses(S^b)
    - Computa a variância entre as classes

$$S^w = \sum_{l=1}^{c} \sum_{i:y_i=l} (x_i - \mu_l)(x_i - \mu_l)^t$$

$$S^b = \sum_{l=1}^{c} N_l(\mu_l - \mu)(\mu_l - \mu)^t$$

LDA

- A função objetivo do LDA é dada por:

$$W_{max} = \underset{\mathbf{W} \in \mathbf{R}^{d \times d}}{\arg\max} \left(\frac{|W^t S^b W|}{|W^t S^w W|} \right)$$

- Com restrição de ortogonalidade: $W^t W = I_d$
- Sendo solucionada pela decomposição em autovalores e autovetores abaixo:

$$\frac{S^b}{S^w} W_{max} = \lambda W_{max}$$

LDA

Demonstração.

assumindo:

$$J(W) = \frac{W^t S^b W}{W^t S^w W}$$

o máximo da função J é dado por $\frac{d}{dW} J(W_{max}) = 0$ então:

$$\frac{d}{dW}\left(\frac{W^t_{max} S^b W_{max}}{W^t_{max} S^w W_{max}}\right) = 0 \;\Rightarrow$$

$$(W^t_{max} S^w W_{max}) \frac{d}{dW}(W^t_{max} S^b W_{max}) - (W^t_{max} S^b W_{max}) \frac{d}{dW}(W^t_{max} S^w W_{max}) = 0 \;\Rightarrow$$

$$(W^t_{max} S^w W_{max}) 2S^b W_{max} - (W^t_{max} S^b W_{max}) 2S^w W_{max} = 0$$

dividindo a expressão por $W^t_{max} S^w W_{max}$:

$$2S^b W_{max} - \frac{W^t_{max} S^b W_{max}}{W^t_{max} S^w W_{max}} 2S^w W_{max} = 0 \;\Rightarrow\; 2S^b W_{max} - J(W_{max}) 2S^w W_{max} = 0 \;\Rightarrow\; J(W_{max}) = \frac{S^b}{S^w}$$

LDA

como $W^t W = I_d$, fazemos:

$$W_{max}^t W_{max} = I_d \;\Rightarrow\; W_{max}^t W_{max} - I_d = 0 \;\Rightarrow$$

$$J(W_{max}) = \frac{S^b}{S^w} W_{max}^t W_{max} - \lambda (W_{max}^t W_{max} - I_d)$$

como $\frac{d}{dW} J(W_{max}) = 0$, temos:

$$\frac{d}{dW}\left[\frac{S^b}{S^w} W_{max}^t W_{max} - \lambda (W_{max}^t W_{max} - I_d)\right] = 0 \;\Rightarrow\; 2\frac{S^b}{S^w} W_{max} - 2\lambda W_{max} = 0 \;\Rightarrow$$

$$\frac{S^b}{S^w} W_{max} = \lambda W_{max}$$

LDA

W_{max}

- $W_{\max} \in R^{dxd}$
- Cada uma de suas colunas, $W_{max}(:,i)$, representa um autovetor de S^b/S^w
- Cada autovetor $W_{max}(:,i)$ está associado a um autovalor $\lambda(i,i)$, cujo o cardinal é diretamente relacionado ao cumprimento dos objetivos do LDA por ele

T_{LDA}

- $T_{LDA} \in R^{dxr} \therefore \quad 1 \leq r \leq d$
- É um subconjunto das colunas de W_{max}
- Agrupa os autovetores de W_{max} correspondentes aos r maiores autovalores em λ
  - Agrupa os r autovetores mais eficazes no cumprimento dos objetivos do LDA

LDA

Ajuste

- Para realizar o cálculo da função objetivo do LDA é preciso que W^tS^wW seja inversível
  - Isto implica que S^w deve ser não-singular (determinante ≠ 0)

$$S^w = S^w + \alpha I_d$$

LDA

Deficiência

- S^b tem no máximo c-1 graus de independência linear
  - Implica que o LDA pode no máximo obter c-1 características significativas

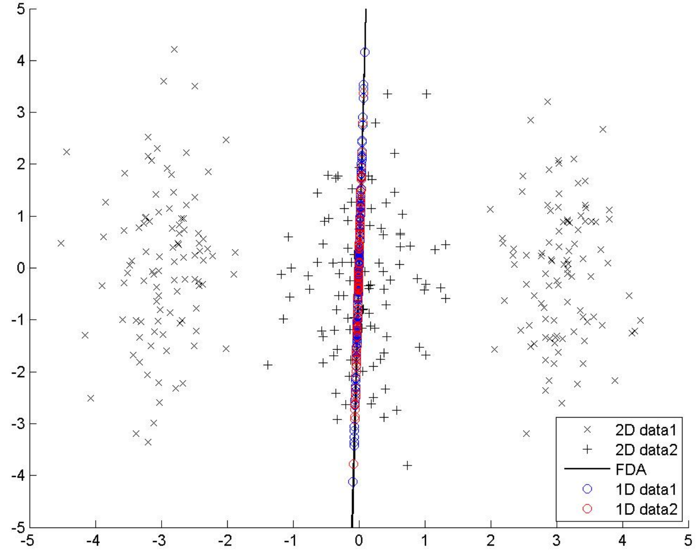
$$1 \leq r \leq (c-1)$$

- Em problemas de duas classes, o único mapeamento possível com o LDA é para um espaço unidimensional

LDA

Deficiência

- LDA não é robusto a distribuição multimodal intraclasse
  - Pois sua modelagem não considera a estrutura local dos padrões
  - Assim, não pode se ajustar de modo a melhor atender classes multimodais



PCA x LDA

PCA

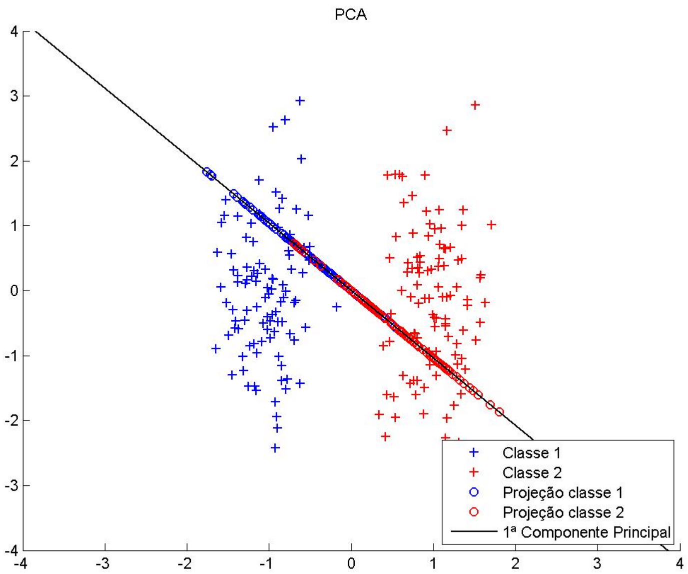
- Não-supervisionada
- Redução de dimensionalidade limitada em $1 \leq r \leq d$
- Não trata multimodalidade
- Falha quando o discriminante não segue a direção das maiores variâncias

LDA

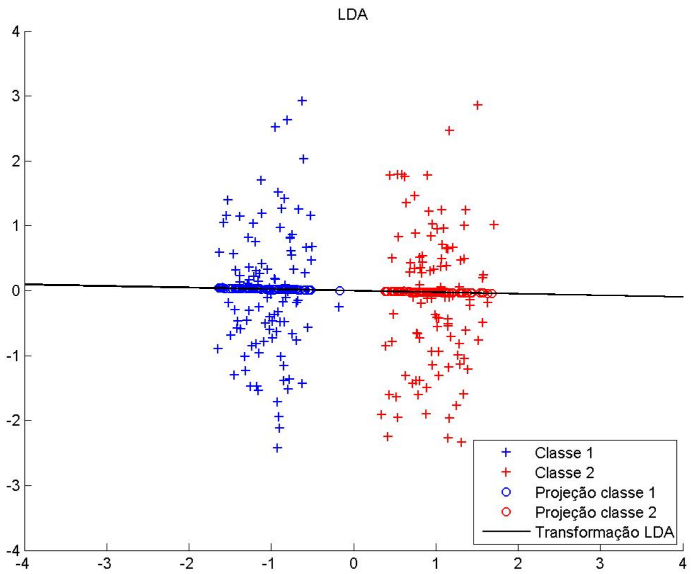
- Supervisionada
- Redução de dimensionalidade limitada em $1 \leq r \leq c-1$
- Não trata multimodalidade

PCA x LDA

PCA

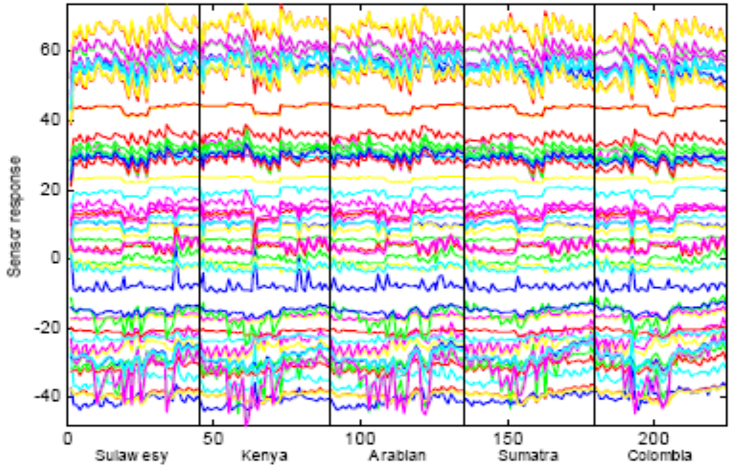


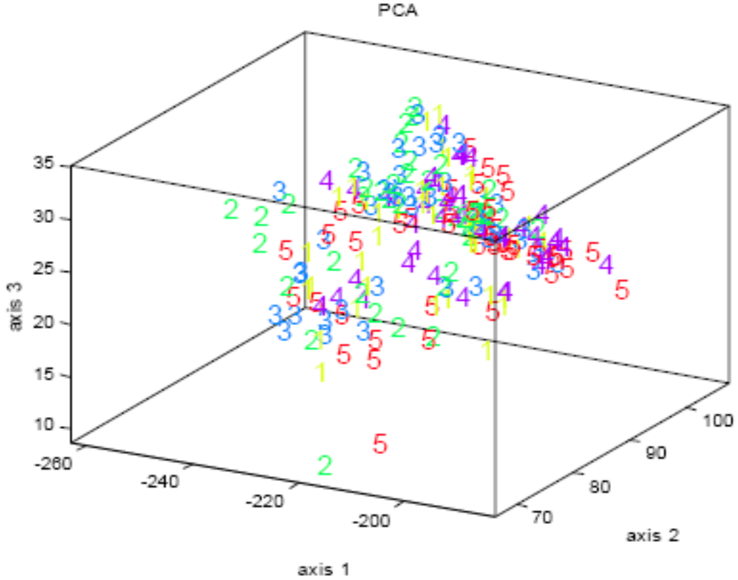
LDA

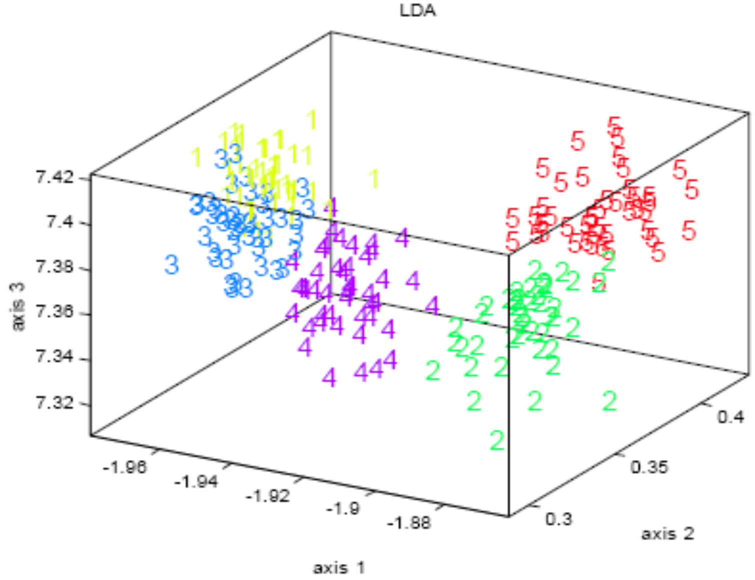


PCA x LDA

- **These figures show the performance of PCA and LDA on an odor recognition problem**
  - Five types of coffee beans were presented to an array of chemical gas sensors
  - For each coffee type, 45 “sniffs” were performed and the response of the gas sensor array was processed in order to obtain a 60-dimensional feature vector
- **Results**
  - From the 3D scatter plots it is clear that LDA outperforms PCA in terms of class discrimination
  - This is one example where the discriminatory information is not aligned with the direction of maximum variance







Implementação - MatLab

```
%Linear Discriminant Analysis
%INPUT:
%       dataset    -   Matriz de padrões. Cada coluna representa um padrão.
%       class      -   Matriz linha com a informação de classe para cada
%                      padrão em dataset
%       nfeatures -    Número de características desejadas para o novo
%                      espaço
%OUTPUT:
%       Tlda       -   Matriz de transformação do LDA.
%                      Esta matriz mapeia o espaço de características
%                      original num novo espaço com menor dimensionalidade
%                      ("nfeatures" dimensões).
function Tlda = LDA(dataset, class, nfeatures)
%lista as classes existentes
classes = unique(class);
% computa a média entre todos os padrões
Mall = mean(dataset,2);
```

Implementação - MatLab

```
% calcula as matrizes de dispersão intraclasse (Sw) e entreclasses (Sb)
Sw = 0;     Sb = 0;
for l = classes
%     encontra os padrões que pertencem a classe "l"
    index = class == l;
%     computa o número de padrões na classe "l"
    nl = sum(index);
%     pega os padrões da classe "l"
    Lsamples = dataset(:,index);
%     calcula a média dos padrões da classe "l"
    Ml = mean(Lsamples,2);
%     monta Sb (matriz de dispersão entreclasses)
    diff = Ml - Mall;
    Sb = Sb + (nl*diff*diff');
%     monta Sw (matriz de dispersão intraclasse)
    for i = 1:nl
        diff = Lsamples(:,i) - Ml;
        Sw = Sw + (diff*diff');
    end
end
```

Implementação - MatLab

```
% AJUSTE - isto é aplicado se Sw for singular
if det(Sw) == 0
    Sw = Sw + 0.00001*eye(size(Sw,1));
end
% decomposição em autovetores (eVectors) e autovalores (eValues)
[eVectors, eValues] = eig(Sw\Sb);% Sw\Sb equivale a inv(Sw)*Sb
% monta Tlda ordenando os autovetores por ordem decrescente de
% significância
[USELESS, index] = sort(sum(eValues,1),'descend');
Tlda = eVectors(:,index(1:nfeatures));
```

Bibliografias

[1] Fisher, R. A.. "The use of multiple measurements in taxonomic problems". Annals Eugen, vol 7, pp 179-188, 1936.

[2] Wen-Hui Yang and Dao-Qing Daí. "Two-Dimensional Maximum Margin Feature Extraction for Face Recognition". IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol 39, nº 4, pp 1002-1012, 2009.

[3] Lin Qi and Enqing Chen and Xiaomin Mu and Ling Guan and Sisi Zhang and Lei Gao. "Recognizing Human Emotional State Based on the 2D-FrFT and FLDA". 2nd International Congress on Image and Signal Processing, pp 1-4, 2009.

[4] Ibrahim, M.T. and Kyan, M. and Ling Guan. "On-line signature verification using global features". Canadian Conference on Electrical and Computer Engineering, pp 682-685, 2009.