Chapter 6

Subspace Methods

Linear Discriminant Analysis

Statistical Methods and Learning in Computer Vision
WS 2012

Claudia Nieuwenhuis
Lehrstuhl für Computer Vision and Pattern Recognition
Fakultät für Informatik
Technische Universität München

Overview

1. **Subspace Methods**

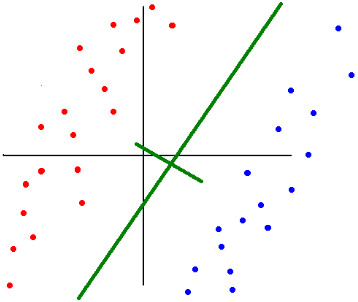
Overview

1. **Subspace Methods**

What is a 'good' subspace?

- Principal Component Analysis
  - Unsupervised, i.e. data is not associated with classes for classification
  - We find an orthogonal coordinate system, i.e. a basis of the subspace
  - The redundancy in the dataset is minimized
- Linear Discriminant Analysis
  - Supervised, i.e. the data points are associated with classes
  - We find an orthogonal coordinate system, i.e. a basis of the subspace
  - The redundancy in the dataset is minimized
- Independent Component Analysis
  - Unupervised, i.e. the data points are not associated with classes
  - We find linear basis vectors, which are not necessarily orthogonal
  - The independencd of different datapoints is maximized

To reduce the dimension of the sample data we can use PCA. If we have classified training data, this method is, however, suboptimal.



Linear Discriminant Analysis

We first start out with only two classes...
We want to find a projection of the data, which satisfies the following goals

- The means of the two classes are as far apart as possible
- The data samples of each class lie as close together as possible

To project a data sample x onto a vector w in 2D, we compute the scalar product $w^T x$.
Let the classes be denoted by C_i with original mean μ_i and Covariance matrix Σ_i before projection.

Within-Class Scatter Matrix

After projection onto direction w the scatter of the samples within each class is measured by

$$\tilde{s}_i^{\,2} = \sum_{y \in \tilde{C}_i} (y - \tilde{\mu}_i)^2 = \sum_{x \in C_i} (w^T x - w^T \mu_i)^2 = w^T \Sigma_i w$$

The sum of the scatter over both classes is given by

$$\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2} = w^T \Sigma_1 w + w^T \Sigma_2 w = w^T (\Sigma_1 + \Sigma_2) w =: w^T S_w w.$$

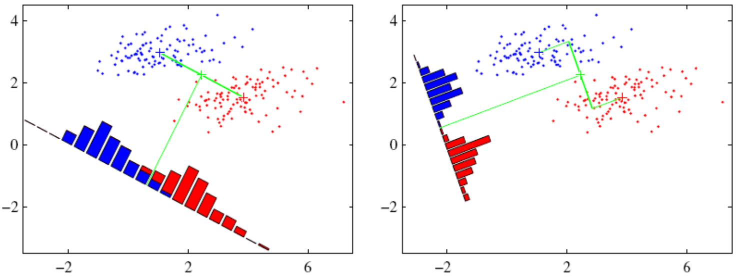
The **within-class scatter matrix** S_w is defined as the sum of the original covariance matrices of the samples in each class.

Between-Class Scatter Matrix

After projection onto direction w the scatter of the means of all classes (between class scatter) is given by

$$\tilde{m} = (\tilde{\mu}_1 - \tilde{\mu}_2)^2 = w^T(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w =: w^T S_B w$$

S_B is called the **between-class scatter matrix**.



Left: projection onto $w = \mu_1 - \mu_2$ is not optimal for separating the samples of the two classes.
Right: projection onto w which minimizes the within class scatter $\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2}$ and maximizes the between class scatter $\tilde{m}$.

Fisher's Discriminant

We want to maximize the scatter between different classes and minimize the scatter within each class. So Fisher proposed to maximize the following function

$$J(w) = \frac{w^T S_B w}{w^T S_W w}$$

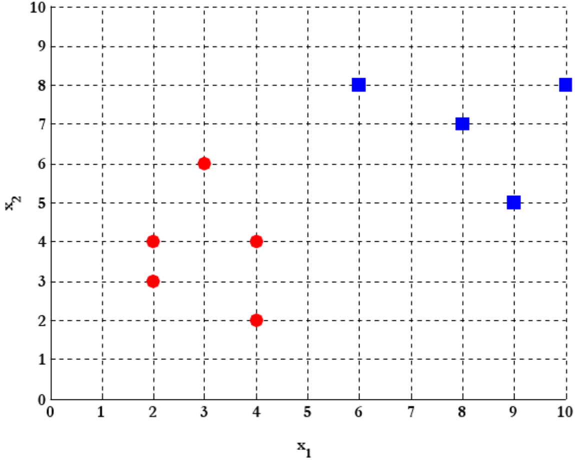
with respect to w. To find the maximum we differentiate and equate to zero.

$$\begin{aligned} \frac{dJ}{dw} &= 0 \Leftrightarrow \\ \underbrace{S_w^{-1} S_B}_{=:A} w &= J(w) w \end{aligned}$$

Provided that S_w is invertible, we have to find the eigenvector with maximum eigenvalue of matrix A to maximize $J(w)$.

LDA Example

- Samples for class $\boldsymbol{\omega}_1$: $\mathbf{X}_1=(x_1,x_2)=\{(4,2),(2,4),(2,3),(3,6),(4,4)\}$
- Sample for class $\boldsymbol{\omega}_2$: $\mathbf{X}_2=(x_1,x_2)=\{(9,10),(6,8),(9,5),(8,7),(10,8)\}$



LDA Example

The LDA projection is obtained by solving the following eigenvalue problem

$$S_W^{-1} S_B w = \lambda w$$

$$\Rightarrow \left| S_W^{-1} S_B - \lambda I \right| = 0$$

$$\Rightarrow \left| \begin{pmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{pmatrix}^{-1} \begin{pmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| = 0$$

$$\Rightarrow \left| \begin{pmatrix} 0.3045 & 0.0166 \\ 0.0166 & 0.1827 \end{pmatrix} \begin{pmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| = 0$$

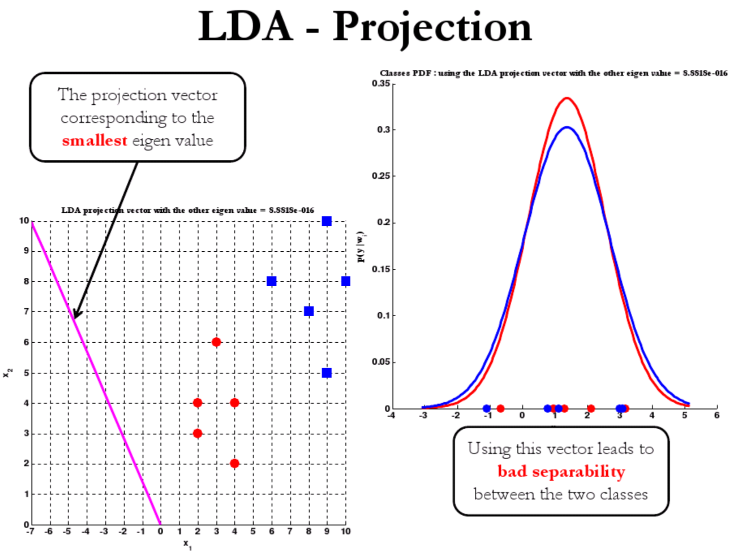
$$\Rightarrow \left| \begin{pmatrix} 9.2213 - \lambda & 6.489 \\ 4.2339 & 2.9794 - \lambda \end{pmatrix} \right|$$

$$= (9.2213 - \lambda)(2.9794 - \lambda) - 6.489 \times 4.2339 = 0$$

$$\Rightarrow \lambda^2 - 12.2007\lambda = 0 \Rightarrow \lambda(\lambda - 12.2007) = 0$$

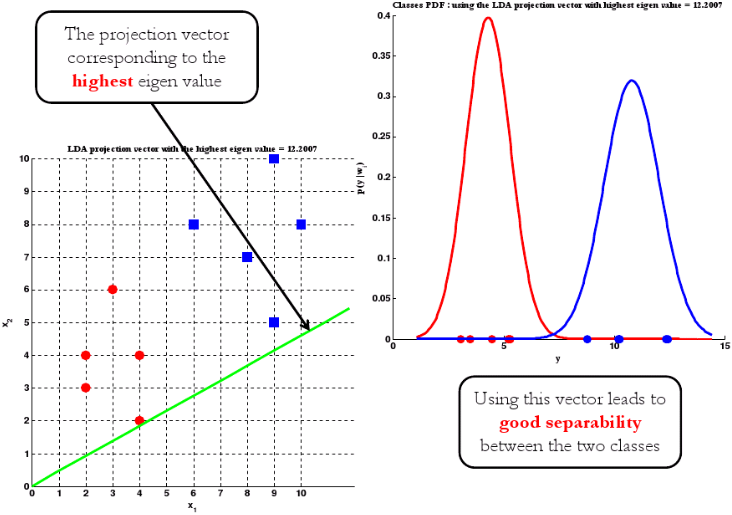
$$\Rightarrow \lambda_1 = 0, \lambda_2 = 12.2007$$

LDA Example


LDA - Projection
The projection vector
corresponding to the
smallest eigen value
LDA projection vector with the other eigen value = 8.8818e-016
Classes PDF : using the LDA projection vector with the other eigen value = 8.8818e-016
p(y|w_i)
Using this vector leads to
bad separability
between the two classes

LDA Example

LDA - Projection

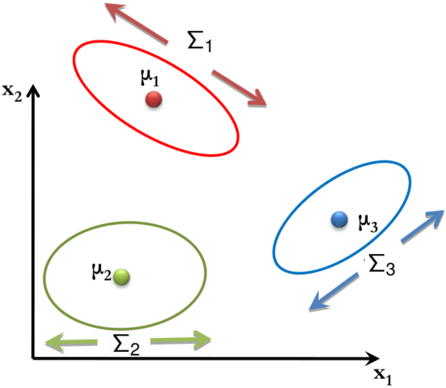


Generalization to multiple classes: within-scatter matrix

For multiple classes $C_1, ..., C_k$ the within-scatter matrix is computed in the following way:

$$S_w = \sum_{i=1}^{k} \Sigma_i$$

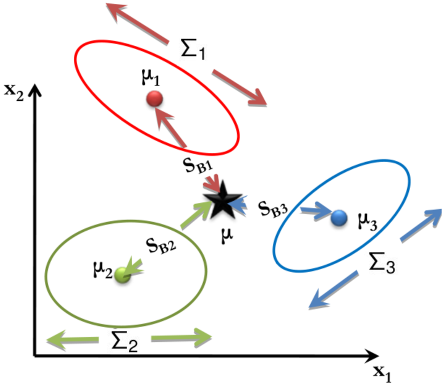
$$\Sigma_i = \frac{1}{n_i - 1} \sum_{x \in C_i} (x - \mu_i)(x - \mu_i)^T, \quad \mu_i = \frac{1}{n_i} \sum_{i \in C_i} x_i$$



Generalization to multiple classes: between-scatter matrix

For multiple classes $C_1, ..., C_k$ the between-scatter matrix is computed in the following way:

$$
\begin{aligned}
S_B &= \sum_{i=1}^{k} S_{B_i} := \sum_{i=1}^{k} N_i(\mu_i - \mu)(\mu_i - \mu)^T \\
\mu &= \frac{1}{N} \sum_{x \in C_1,\dots,C_k} x
\end{aligned}
$$



Linear Discriminant Analysis

Claudia Nieuwenhuis

In multiple dimensions, we now want to find the matrix W containing $k-1$ columns, which maximizes the following ratio

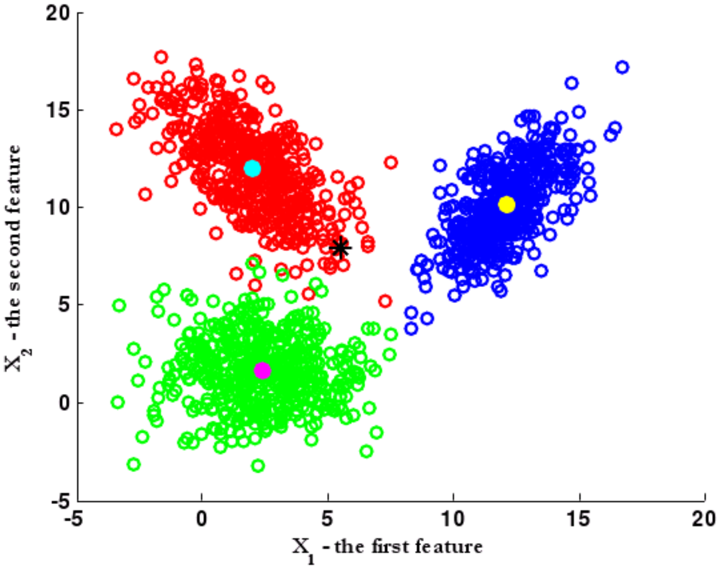
$$J(W) = \frac{det(W^T S_B W)}{det(W^T S_w W)}$$

Since the projection is no longer a scalar, the determinant is used. In multiple dimensions, the optimization leads to the same eigenvalue equation

$$\underbrace{S_w^{-1} S_B}_{=:A} W = J(w) W.$$

Hence, the matrix W consists of the $k-1$ eigenvectors with largest eigenvalues of A in its columns.

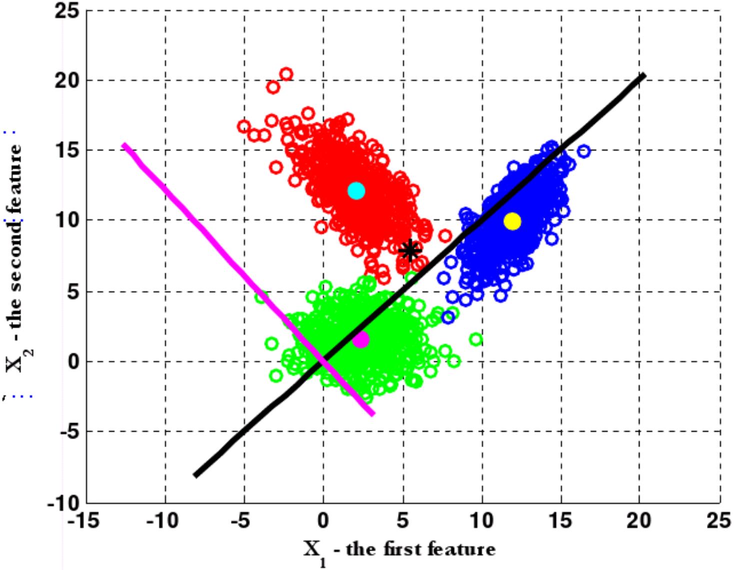
LDA Example



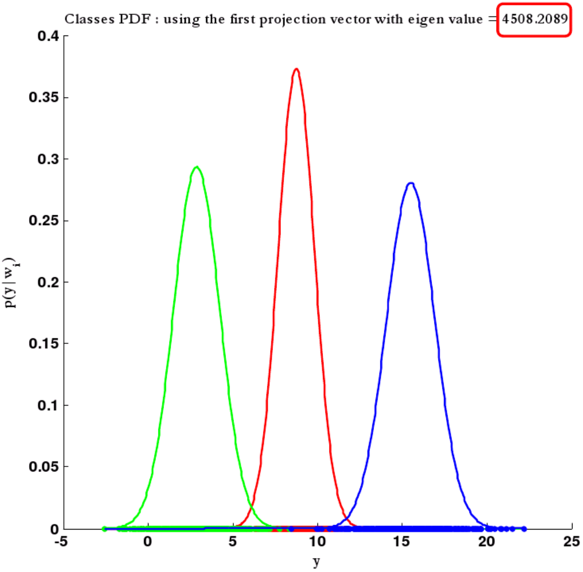
20
15
10
5
0
-5
-5
0
5
10
15
20
X_2 - the second feature
X_1 - the first feature

LDA Example

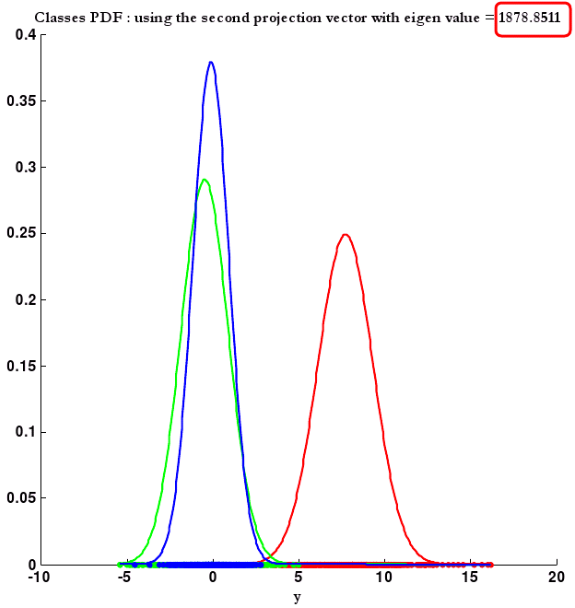
The LDA projection is obtained by solving the following eigenvalue problem



LDA Example


Classes PDF : using the first projection vector with eigen value = 4508.2089
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
p(y|w_i)
-5
0
5
10
15
20
25
y

LDA Example


Classes PDF : using the second projection vector with eigen value = 1878.8511
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-10
-5
0
5
10
15
20
y