Prof. Marios Savvides

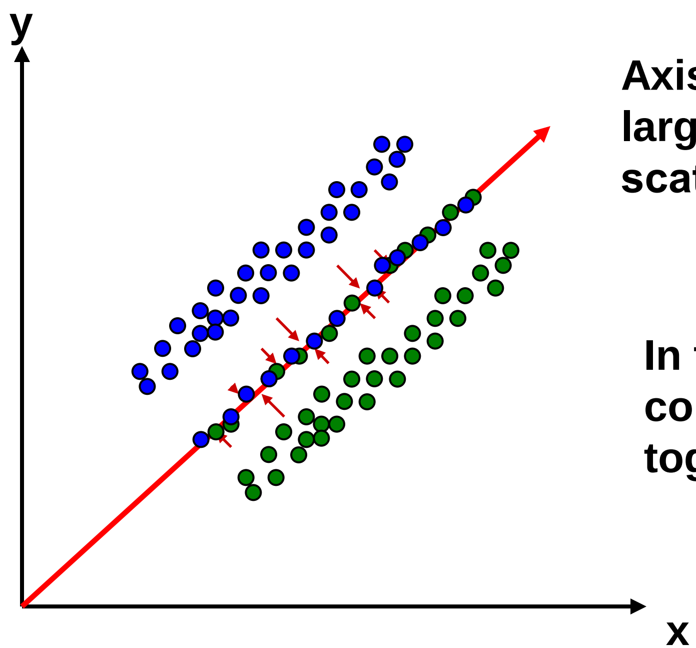
Pattern Recognition Theory

Lecture 8 : Linear Discriminant Analysis

Electrical & Computer ENGINEERING

What is PCA?
What are we trying to do?

- We want to find projections of data (i.e. direction vectors that we can project the data on to) that describe the maximum variation.

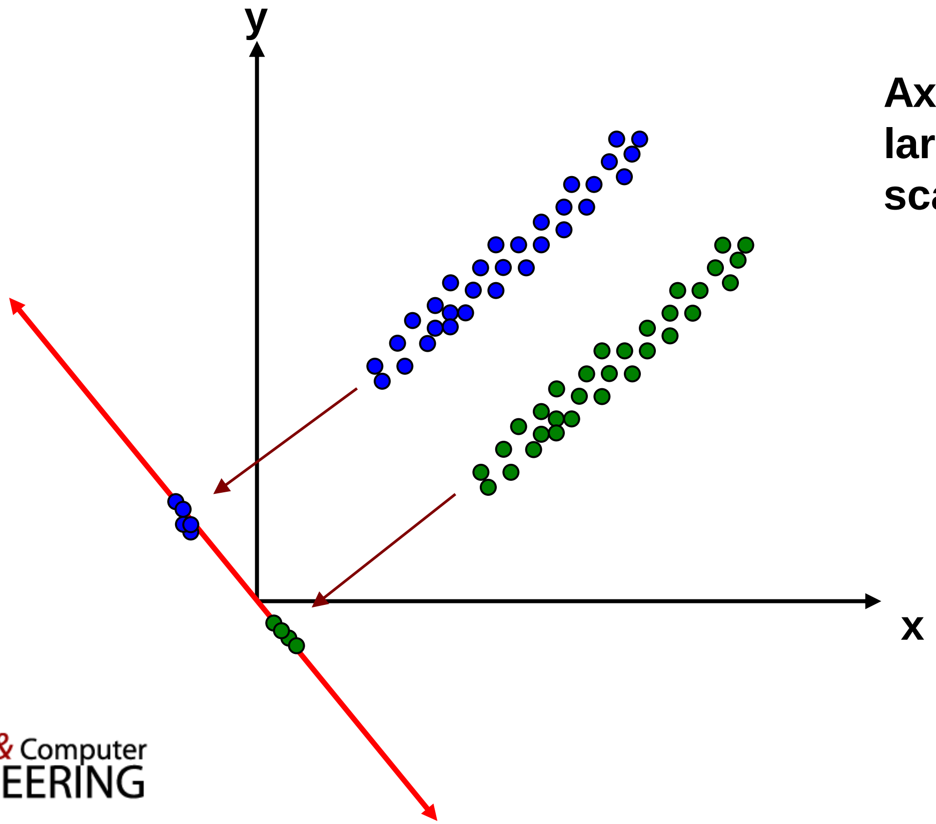


Axis that describes the largest variation (or scatter)…

In this case the projection vector completely smears the two classes together, making them in-separable

What is LDA?
What are we trying to do?

- We want to find projections that separate the classes. (Assume unimodal Gaussian modes – maximize distance between two means and minimize variance =>will lead to minimize overall probability of error



Case 1: Simple 2 Class Problem

- We want to maximize the distance between the projected means:

e.g. maximize $|(\tilde{\mu}_1 - \tilde{\mu}_2)|^2$

Where $\tilde{\mu}_1$ is the projected mean $\boldsymbol{\mu}_1$ of class onto LDA direction vector $\mathbf{w}$, i.e.

$$\tilde{\mu}_1 = \mathbf{w}^{\mathbf{T}}\boldsymbol{\mu}_1$$

and for class 2: $\tilde{\mu}_2 = \mathbf{w}^{\mathbf{T}}\boldsymbol{\mu}_2$ thus

$$|(\tilde{\mu}_1 - \tilde{\mu}_2)|^2 = |(\mathbf{w}^{\mathbf{T}}\boldsymbol{\mu}_1 - \mathbf{w}^{\mathbf{T}}\boldsymbol{\mu}_2)|^2$$

$$= \mathbf{w}^{\mathbf{T}}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathbf{T}}\mathbf{w}$$

$$= \mathbf{w}^{\mathbf{T}}\mathbf{S}_{\mathbf{B}}\mathbf{w}$$

Between Class Scatter Matrix S_B

$$(\tilde{\mu}_1 - \tilde{\mu}_2)^2 = (\mathbf{w}^{\mathbf{T}}\boldsymbol{\mu}_1 - \mathbf{w}^{\mathbf{T}}\boldsymbol{\mu}_2)^2$$
$$= \mathbf{w}^{\mathbf{T}}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathbf{T}}\mathbf{w}$$
$$= \mathbf{w}^{\mathbf{T}}\mathbf{S}_{\mathbf{B}}\mathbf{w}$$

We want to maximize $w^T S_B w$ where S_B is the between class scatter matrix defined as:

$$\mathbf{S}_{\mathbf{B}} = (\boldsymbol{\mu}_{\mathbf{1}} - \boldsymbol{\mu}_{\mathbf{2}})(\boldsymbol{\mu}_{\mathbf{1}} - \boldsymbol{\mu}_{\mathbf{2}})^{\mathbf{T}}$$

NOTE: S_B is rank 1. This will be useful later on to find closed form solution for 2-class LDA

We also want to minimize….

- The variance or scatter of the projected samples from each class (i.e. we want to make each class more compact or closer to its mean). The scatter from class 1 defined as s_1 is given as

$$\tilde{s}_1^{\,2} = \sum_{i=1}^{N_1} (\tilde{x}_i - \tilde{\mu}_1)^2$$

- Thus we want to minimize the scatter of class 1 and class 2 in projected space, i.e.

minimize the total scatter $\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2}$

Fisher Linear Discriminant Criterion Function

- Objective [1]-We want to ***maximize*** the between class scatter defined as: $|(\tilde{\mu}_1 - \tilde{\mu}_2)|^2$
- [2]- We want to ***minimize*** the within-class scatter.

$$\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2}$$

- Thus we define our objective function *J*(**w**) as the following ratio that we want to ***maximize*** in order to achieve [1] and [2]:

$$J(\mathbf{w}) = \frac{|(\tilde{\mu}_1 - \tilde{\mu}_2)|^2}{\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2}}$$

Scatter s_1 of Class 1

- Thus we want to find the vector **w** that maximizes J(**w**).
- Lets expand on scatter

$$\tilde{s}_1^{\,2} = \sum_{i=1}^{N_1} (\tilde{x}_i - \tilde{\mu}_1)^2$$

$$= \sum_{i=1}^{N_1} (\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}_1)^2$$

$$= \sum_{i=1}^{N_1} \mathbf{w}^{\mathrm{T}}(\mathbf{x}_i - \boldsymbol{\mu}_1)(\mathbf{x}_i - \boldsymbol{\mu}_1)^{\mathrm{T}}\mathbf{w}$$

$$= \mathbf{w}^{\mathrm{T}}\mathbf{S}_1\mathbf{w}$$

Same for Scatter s_2 of Class 2

$$\tilde{s}_2^{\,2} = \sum_{i=1}^{N_2} (\tilde{x}_i - \tilde{\mu}_2)^2$$

$$= \sum_{i=1}^{N_2} (\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}_2)^2$$

$$= \sum_{i=1}^{N_2} \mathbf{w}^{\mathrm{T}}(\mathbf{x}_i - \boldsymbol{\mu}_2)(\mathbf{x}_i - \boldsymbol{\mu}_2)^{\mathrm{T}}\mathbf{w}$$

$$= \mathbf{w}^{\mathrm{T}}\mathbf{S}_2\mathbf{w}$$

Total Within-Class Scatter Matrix

- We want to minimize total *within*-class scatter. i.e.

$$\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2}$$

- Which is equivalent to $\mathbf{w}^{\mathbf{T}}\mathbf{S}_{\mathrm{w}}\mathbf{w}$

$$\mathbf{S_w} = \sum_{i=1}^{C}\sum_{j=1}^{Ni} (\mathbf{x}_j - \boldsymbol{\mu}_i)(\mathbf{x}_j - \boldsymbol{\mu}_i)^{\mathbf{T}}$$

C=2, N_i=No. of images in *i*th class

Solving LDA

- Maximize $J(\mathbf{w}) = \dfrac{\mathbf{w}^{\mathrm{T}} \mathbf{S_B} \mathbf{w}}{\mathbf{w}^{\mathrm{T}} \mathbf{S_W} \mathbf{w}}$
- We need to find the optimal **w** which will maximize the above ratio (or quotient).
- So what do we do now?

  Take derivative and solve for **w**

Solving LDA

$$J(\mathbf{w}) = \frac{\mathbf{w}^{\mathbf{T}}\mathbf{S}_{\mathbf{B}}\mathbf{w}}{\mathbf{w}^{\mathbf{T}}\mathbf{S}_{\mathbf{W}}\mathbf{w}}$$

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \frac{2\mathbf{w}^{\mathbf{T}}\mathbf{S}_{\mathbf{W}}\mathbf{w}\mathbf{S}_{\mathbf{B}}\mathbf{w} - 2\mathbf{w}^{\mathbf{T}}\mathbf{S}_{\mathbf{B}}\mathbf{w}\mathbf{S}_{\mathbf{W}}\mathbf{w}}{(\mathbf{w}^{\mathbf{T}}\mathbf{S}_{\mathbf{W}}\mathbf{w})^2} = 0$$

$$=> \frac{\mathbf{w}^{\mathbf{T}}\mathbf{S}_{\mathbf{W}}\mathbf{w}\mathbf{S}_{\mathbf{B}}\mathbf{w} - \mathbf{w}^{\mathbf{T}}\mathbf{S}_{\mathbf{B}}\mathbf{w}\mathbf{S}_{\mathbf{W}}\mathbf{w}}{\mathbf{w}^{\mathbf{T}}\mathbf{S}_{\mathbf{W}}\mathbf{w}} = 0$$

$$=> \mathbf{S}_{\mathbf{B}}\mathbf{w} - J(\mathbf{w})\mathbf{S}_{\mathbf{W}}\mathbf{w} = 0$$

Solving LDA

$$\mathbf{S_B w} - J(\mathbf{w})\mathbf{S_W w} = 0$$

$$\mathbf{S_B w} - \lambda \mathbf{S_W w} = 0$$

$$\mathbf{S_B w} = \lambda \mathbf{S_W w}$$

Generalized Eigenvalue problem

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

If S_w is non-singular and invertible

We want to maximize J(w) thus we want the eigenvector w with the largest eigenvalue!