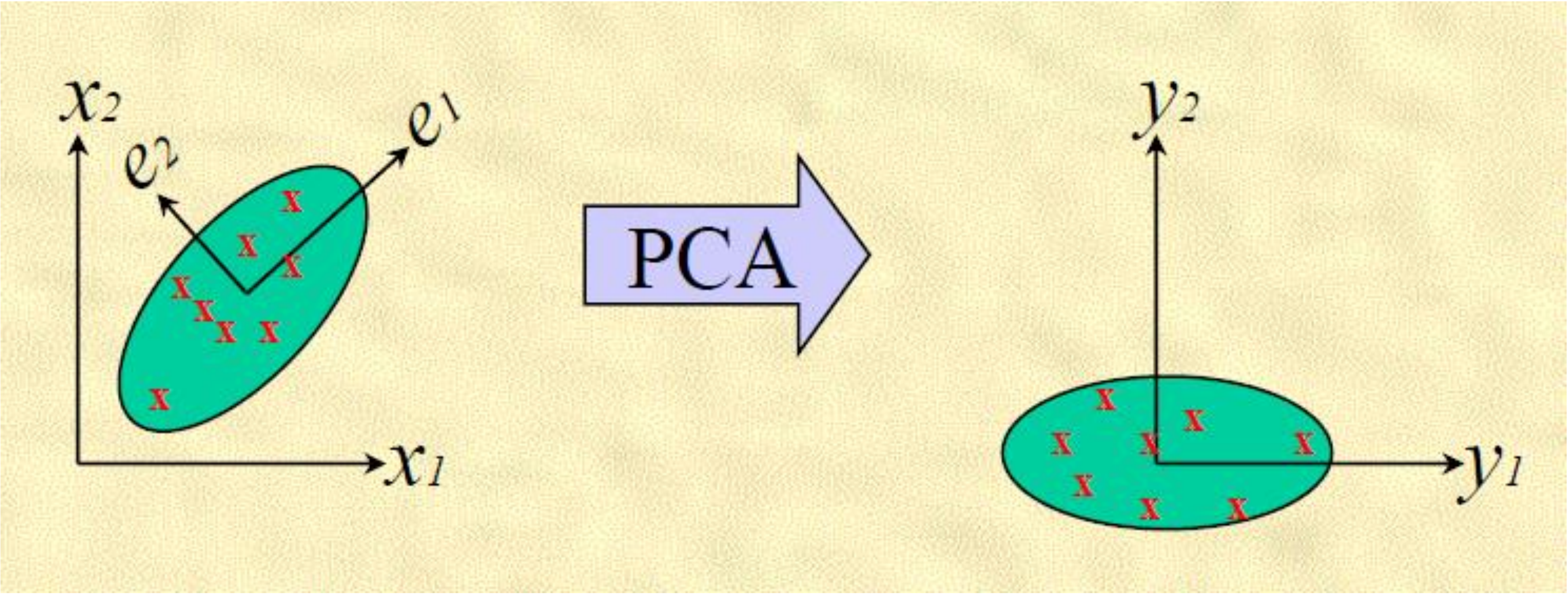
Advanced Digital Image Processing

Week 16

Khawar Khurshid, 2012

Principal Component Analysis



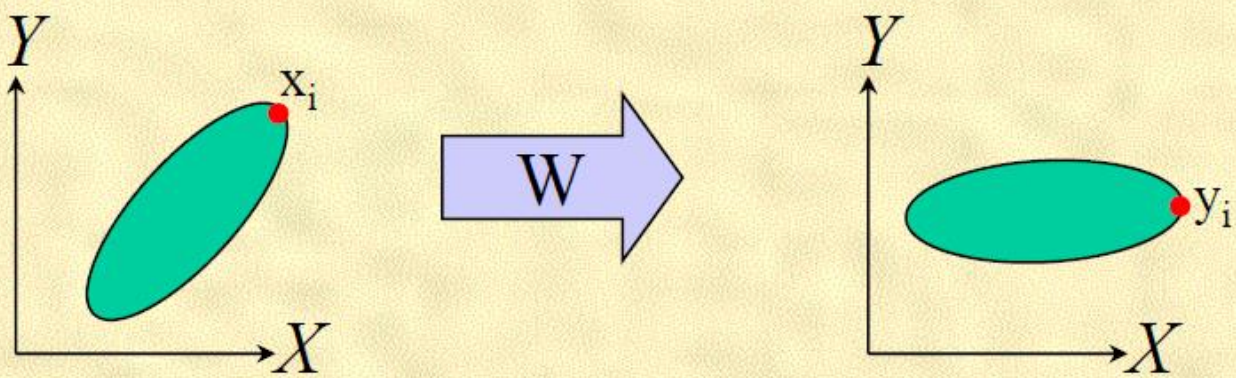
Rotate the data so that its primary axes lie along the axes of the coordinate space and move it so that its center of mass lies on the origin.

Principal Component Analysis

Problem formulation

- Input: $x=[x_1|\ldots|x_N]_{d'N}$ points in d-dimensional space
- Look for: W $d'm$ projection matrix $(m \leq d)$
- S.t. : $y=[y_1|\ldots|y_N]_{m'N} = W^T [x_1|\ldots|x_N]$...

  ...And correlation is minimized



Principal Component Analysis

- Define the covariance (scatter) matrix of the input samples as :

$$\boxed{Cov(x) = S_T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T}$$

(where μ is the sample mean)

- The matrix Cov is symmetric and of dimension d′d.
- The diagonal contains the variance of each parameter (i.e. element Cov_{ii} is the variance in the i'th direction).
- Each element Cov_{ij} is the co-variance between the two directions i and j, or how correlated are they (i.e. a value of zero indicates that the two dimensions are uncorrelated).

Principal Component Analysis

How do we find such W ?

$$\lambda_i W_i = Cov(x) W_i$$

Therefore :

Choose W_{opt} to be the eigenvectors matrix:

$$W_{opt} = [w_1 | \dots | w_d]$$

Where $\{w_i | i=1,\dots,d\}$ is the set of the d-dimensional eigenvectors of $Cov(x)$!

Principal Component Analysis

- To find a more convenient coordinate system one needs to :

Calculate mean sample μ → Subtract it from all samples x_i → Calculate Covariance matrix for resulting samples → Find the set of eigenvectors for the covariance matrix

↓

Create W_{opt}, the projection matrix, by taking as columns the eigenvectors calculated !

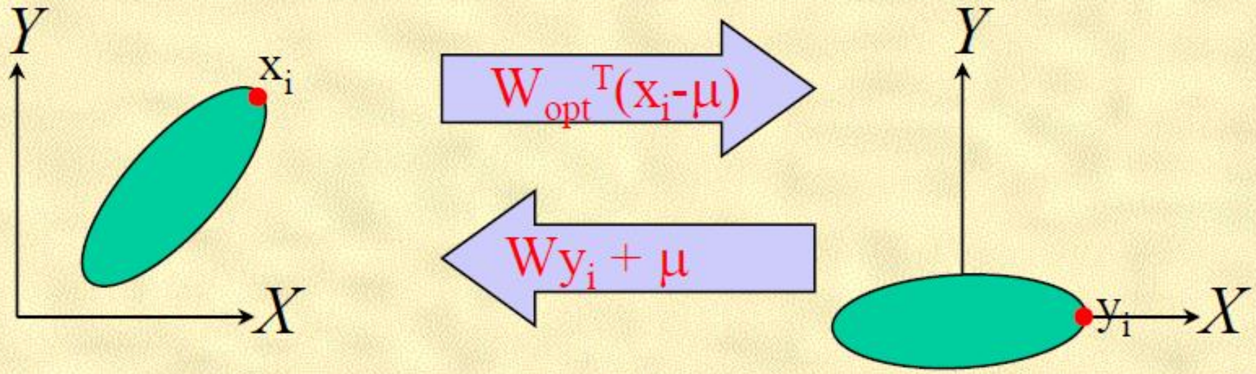
Principal Component Analysis

- Now we have that any point x_i can be projected to an appropriate point y_i by :

$$y_i = W_{opt}^T(x_i - \mu)$$

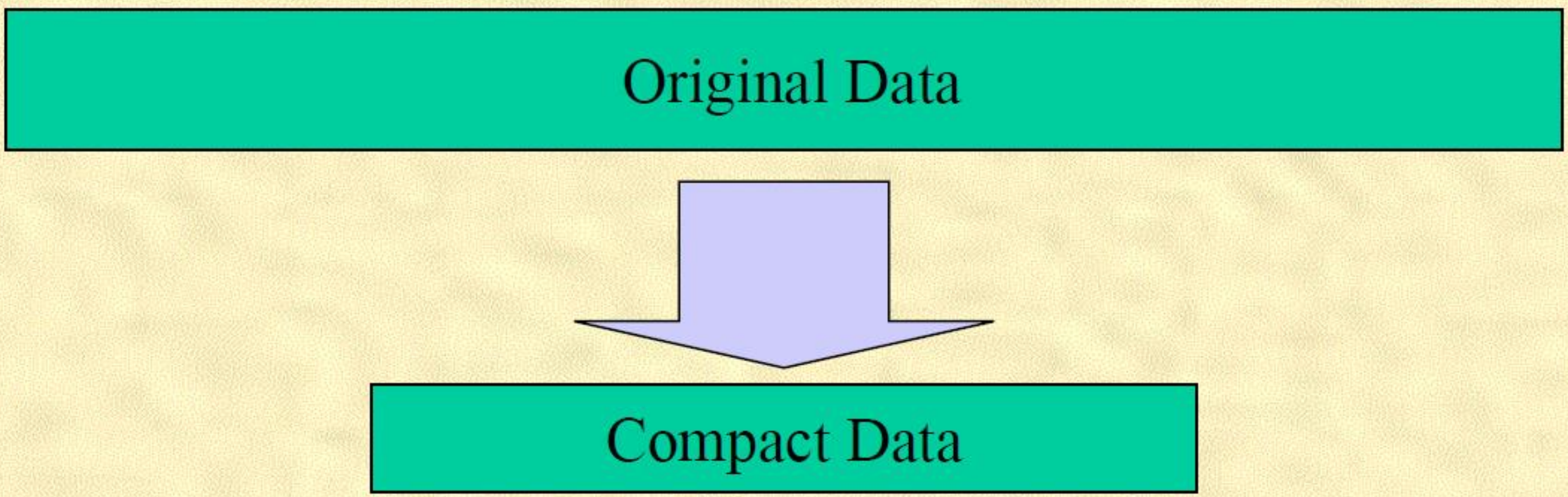
- and conversely (since $W^{-1} = W^T$)

$$Wy_i + \mu = x_i$$

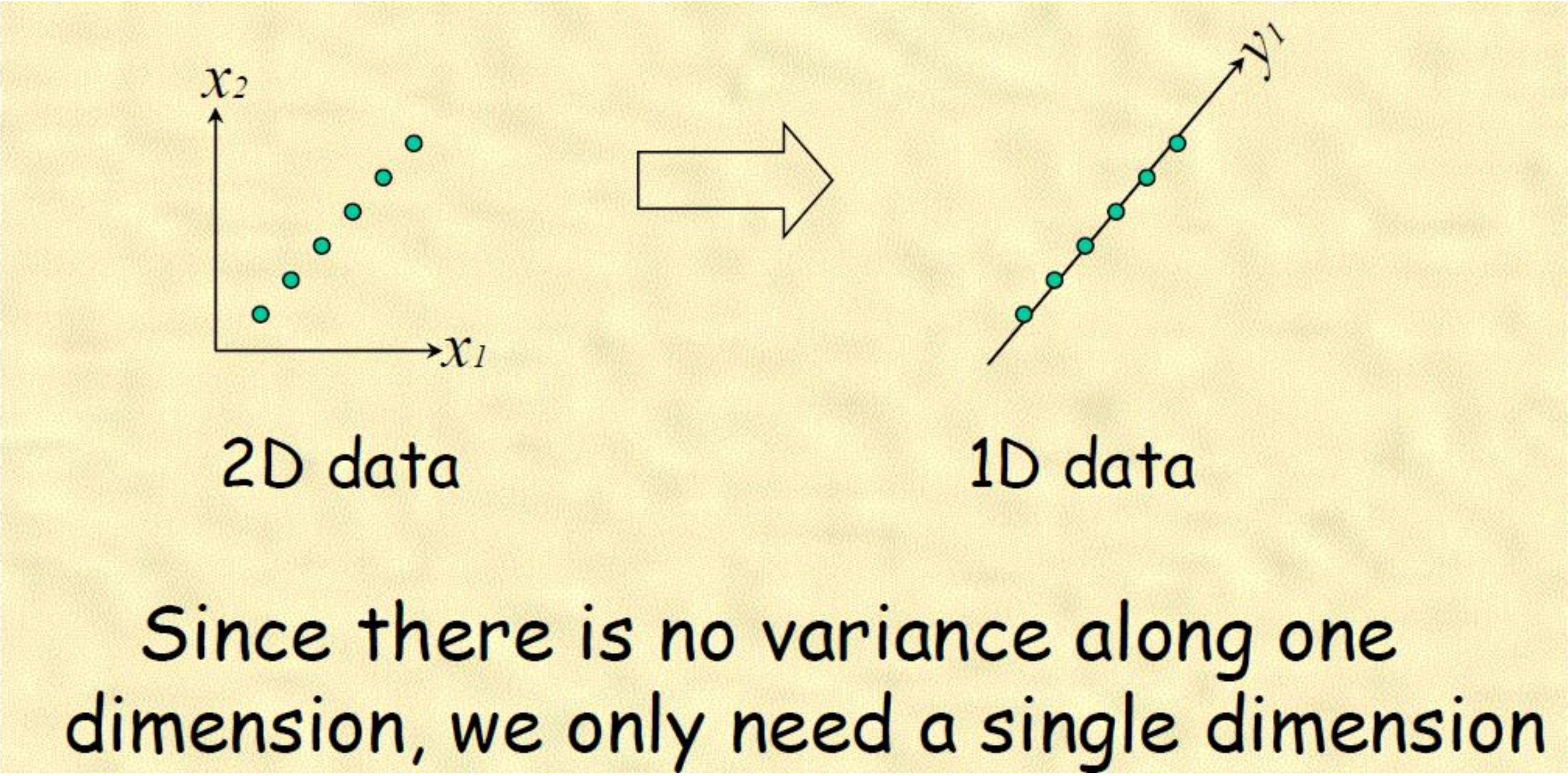


Principal Component Analysis

Reduce space dimensionality with minimum loss of description information.



Principal Component Analysis



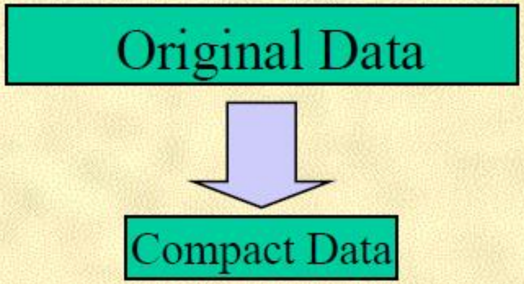
Principal Component Analysis

- Each eigenvalue represents the the total variance in its dimension.
- So…
- Throwing away the least significant eigenvectors in W_{opt} means throwing away the least significant variance information !

Principal Component Analysis

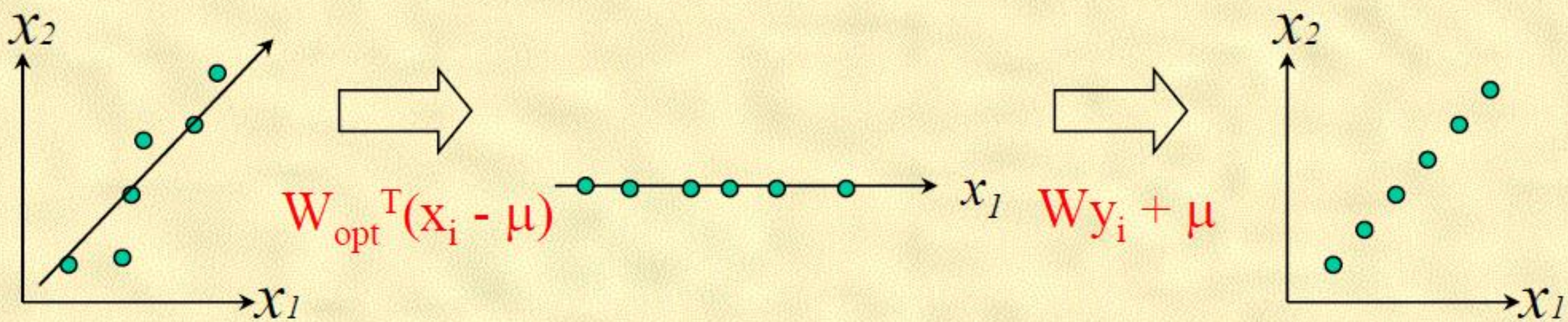
- Sort the d columns of the projection matrix W_{opt} in descending order of appropriate eigenvalues.
- Select the first m columns thus creating a new projection matrix of dimension d´m

This will now be a projection from a d-dimensional space to an m-dimentional space (m < d) !



Principal Component Analysis

- Sample points can still be projected via the new m′d projection matrix W_{opt} and can still be reconstructed, but some information will be lost.



2D data 1D data 2D data

Principal Component Analysis

- It can be shown that the mean square error between x_i and its reconstruction using only m principle eigenvectors is given by the expression :

$$\boxed{\sum_{j=1}^{N} \lambda_j - \sum_{j=1}^{m} \lambda_j = \sum_{j=m+1}^{N} \lambda_j}$$

Fisher Linear Discriminant

Fisher Linear Discriminant

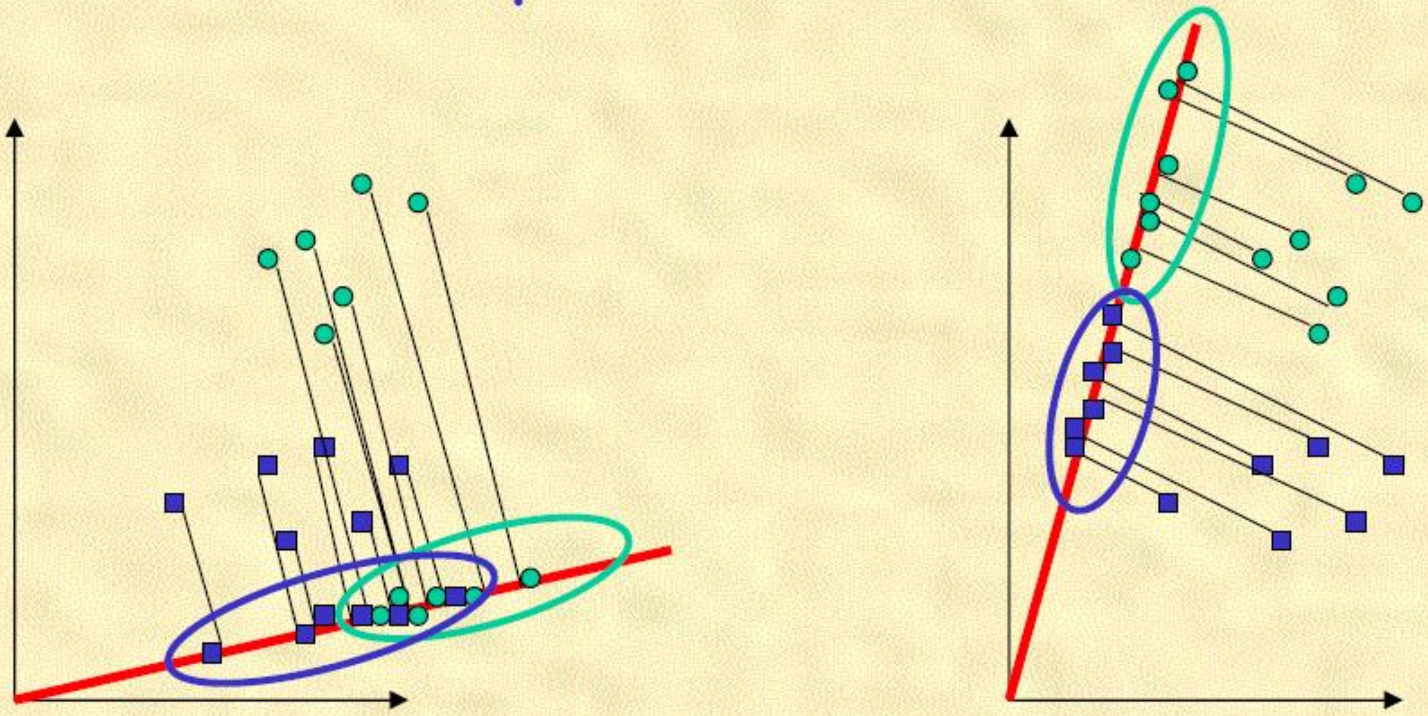
The objective of LDA is to perform dimensionality reduction …

So what, PCA does this!!

However, we want to preserve as much of the class discriminatory information as possible.

Fisher Linear Discriminant

- Objective: Find a projection which separates data clusters



Poor separation

Good separation

Fisher Linear Discriminant

- In order to find a good projection vector, we need to define a measure of separation
- The mean vector of each class in x-space and y-space is

$$\mu_i = \frac{1}{N_i}\sum_{x\in\omega_i} x \text{ and } \tilde{\mu}_i = \frac{1}{N_i}\sum_{y\in\omega_i} y = \frac{1}{N_i}\sum_{x\in\omega_i} w^T x = w^T \mu_i$$

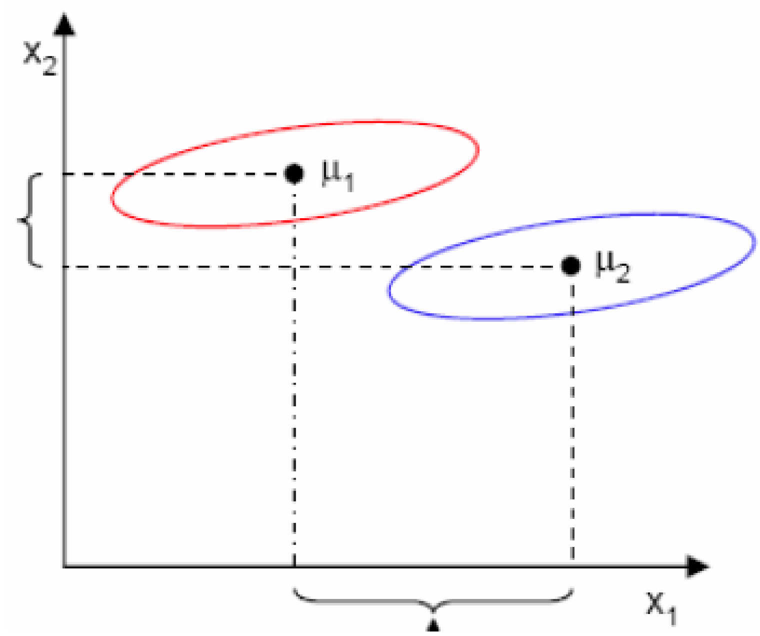
- We could then choose the distance between the projected means as our objective function

$$J(w) = |\tilde{\mu}_1 - \tilde{\mu}_2| = |w^T(\mu_1 - \mu_2)|$$

Fisher Linear Discriminant

- However, the distance between the projected means is not a very good measure since it does not take into account the standard deviation within the classes.

This axis yields better class separability



This axis has a larger distance between means

Fisher Linear Discriminant

- The solution proposed by Fisher is to maximize a function that represents the difference between the means, normalized by a measure of the within-class variability, or the so-called *scatter*.

- For each class we define the scatter, an equivalent of the variance, as;

$$\tilde{s}_i^{\,2} = \sum_{y \in \omega_i} (y - \tilde{\mu}_i)^2$$

- $\tilde{s}_i^{\,2}$ measures the variability within class $\boldsymbol{\omega}_i$ after projecting it on the y-space.

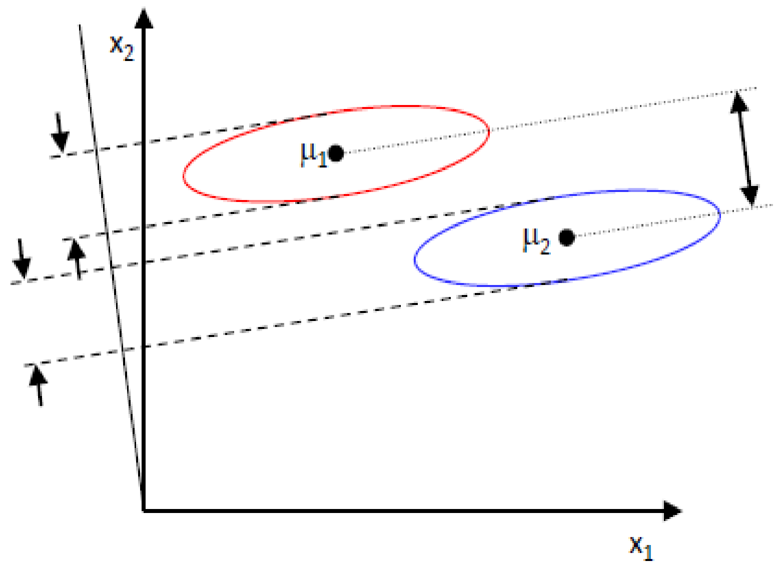
- Thus $\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2}$ measures the variability within the two classes at hand after projection, hence it is called *within-class scatter* of the projected samples.

Fisher Linear Discriminant

- The Fisher linear discriminant is defined as the linear function $w^T x$ that maximizes the criterion function

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

- Therefore, we are looking for a projection where examples from the same class are projected very close to each other and, at the same time, the projected means are as farther apart as possible



Fisher Linear Discriminant

- In order to find the optimum projection $\mathbf{w}^*$, we need to express *J(w)* as an explicit function of **w**.
- We will define a measure of the scatter in multivariate feature space **x** which are denoted as *scatter matrices*;

$$S_i = \sum_{x \in \omega_i} (x - \mu_i)(x - \mu_i)^T$$

$$S_w = S_1 + S_2$$

- Where $\mathbf{S}_i$ is the covariance matrix of class $\boldsymbol{\omega}_i$, and $\mathbf{S}_w$ is called the ***within-class scatter matrix***.

Fisher Linear Discriminant

- Now, the scatter of the projection **y** can then be expressed as a function of the scatter matrix in feature space **x**.

$$\tilde{s}_i^2 = \sum_{y \in \omega_i} (y - \tilde{\mu}_i)^2 = \sum_{x \in \omega_i} \left(w^T x - w^T \mu_i\right)^2$$

$$= \sum_{x \in \omega_i} w^T (x - \mu_i)(x - \mu_i)^T w$$

$$= w^T S_i w$$

$$\tilde{s}_1^2 + \tilde{s}_2^2 = w^T S_1 w + w^T S_2 w = w^T (S_1 + S_2) w = w^T S_W w = \tilde{S}_W$$

Where $\tilde{S}_W$ is the within-class scatter matrix of the projected samples **y**.

Fisher Linear Discriminant

- Similarly, the difference between the projected means (in y-space) can be expressed in terms of the means in the original feature space (x-space).

$$(\tilde{\mu}_1 - \tilde{\mu}_2)^2 = \left(w^T \mu_1 - w^T \mu_2\right)^2$$

$$= w^T \underbrace{(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T}_{S_B} w$$

$$= w^T S_B w = \tilde{S}_B$$

- The matrix $\mathbf{S_B}$ is called the ***b****etween-class scatter* of the original samples/feature vectors, while $\tilde{S}_B$ is the between-class scatter of the projected samples **y**.

Fisher Linear Discriminant

- We can finally express the Fisher criterion in terms of $\mathbf{S_W}$ and $\mathbf{S_B}$ as:

$$J(w) = \frac{\left|\tilde{\mu}_1 - \tilde{\mu}_2\right|^2}{\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2}} = \frac{w^T S_B w}{w^T S_W w}$$

- Hence $J(w)$ is a measure of the difference between class means (encoded in the between-class scatter matrix) normalized by a measure of the within-class scatter matrix.

Fisher Linear Discriminant

- To find the maximum of $J(w)$, we differentiate and equate to zero.

$$\frac{d}{dw}J(w)=\frac{d}{dw}\left(\frac{w^T S_B w}{w^T S_W w}\right)=0$$

$$\Rightarrow \left(w^T S_W w\right)\frac{d}{dw}\left(w^T S_B w\right)-\left(w^T S_B w\right)\frac{d}{dw}\left(w^T S_W w\right)=0$$

$$\Rightarrow \left(w^T S_W w\right)2S_B w-\left(w^T S_B w\right)2S_W w=0$$

$$Dividing\ by\ 2w^T S_W w:$$

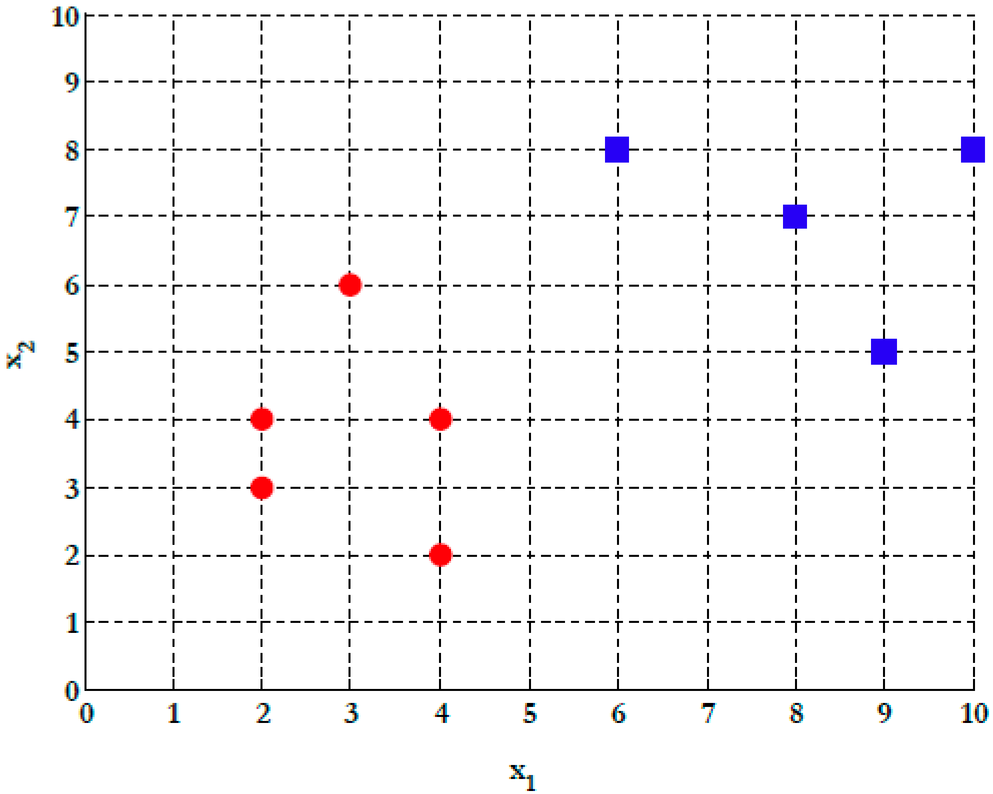
$$\Rightarrow \left(\frac{w^T S_W w}{w^T S_W w}\right)S_B w-\left(\frac{w^T S_B w}{w^T S_W w}\right)S_W w=0$$

$$\Rightarrow S_B w - \lambda S_W w=0$$

$$\Rightarrow S_W^{-1} S_B w - \lambda w=0$$

Fisher Linear Discriminant

- Compute the Linear Discriminant projection for the following two-dimensional dataset.
  - Samples for class ω_1 : $\mathbf{X}_1=(x_1,x_2)=\{(4,2),(2,4),(2,3),(3,6),(4,4)\}$
  - Sample for class ω_2 : $\mathbf{X}_2=(x_1,x_2)=\{(9,10),(6,8),(9,5),(8,7),(10,8)\}$




```
% samples for class 1
X1 = [4,2;
      2,4;
      2,3;
      3,6;
      4,4];


% samples for class 2
X2 = [9,10;
      6,8;
      9,5;
      8,7;
      10,8];
```

Fisher Linear Discriminant

- The classes mean are :

$$\mu_1 = \frac{1}{N_1}\sum_{x\in\omega_1} x = \frac{1}{5}\left[\begin{pmatrix}4\\2\end{pmatrix}+\begin{pmatrix}2\\4\end{pmatrix}+\begin{pmatrix}2\\3\end{pmatrix}+\begin{pmatrix}3\\6\end{pmatrix}+\begin{pmatrix}4\\4\end{pmatrix}\right]=\begin{pmatrix}3\\3.8\end{pmatrix}$$

$$\mu_2 = \frac{1}{N_2}\sum_{x\in\omega_2} x = \frac{1}{5}\left[\begin{pmatrix}9\\10\end{pmatrix}+\begin{pmatrix}6\\8\end{pmatrix}+\begin{pmatrix}9\\5\end{pmatrix}+\begin{pmatrix}8\\7\end{pmatrix}+\begin{pmatrix}10\\8\end{pmatrix}\right]=\begin{pmatrix}8.4\\7.6\end{pmatrix}$$

```
% class means
Mu1 = mean(X1)';
Mu2 = mean(X2)';
```

Fisher Linear Discriminant

- Between-class scatter matrix:

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

$$= \left[\begin{pmatrix} 3 \\ 3.8 \end{pmatrix} - \begin{pmatrix} 8.4 \\ 7.6 \end{pmatrix}\right]\left[\begin{pmatrix} 3 \\ 3.8 \end{pmatrix} - \begin{pmatrix} 8.4 \\ 7.6 \end{pmatrix}\right]^T$$

$$= \begin{pmatrix} -5.4 \\ -3.8 \end{pmatrix}\begin{pmatrix} -5.4 & -3.8 \end{pmatrix}$$

$$= \begin{pmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{pmatrix}$$

```
% between-class scatter matrix
SB = (Mu1-Mu2)*(Mu1-Mu2)';
```

Fisher Linear Discriminant

- Covariance matrix of the first class:

$$S_1 = \sum_{x \in \omega_1} (x - \mu_1)(x - \mu_1)^T = \left[\begin{pmatrix} 4 \\ 2 \end{pmatrix} - \begin{pmatrix} 3 \\ 3.8 \end{pmatrix}\right]^2 + \left[\begin{pmatrix} 2 \\ 4 \end{pmatrix} - \begin{pmatrix} 3 \\ 3.8 \end{pmatrix}\right]^2$$

$$+ \left[\begin{pmatrix} 2 \\ 3 \end{pmatrix} - \begin{pmatrix} 3 \\ 3.8 \end{pmatrix}\right]^2 + \left[\begin{pmatrix} 3 \\ 6 \end{pmatrix} - \begin{pmatrix} 3 \\ 3.8 \end{pmatrix}\right]^2 + \left[\begin{pmatrix} 4 \\ 4 \end{pmatrix} - \begin{pmatrix} 3 \\ 3.8 \end{pmatrix}\right]^2$$

$$= \begin{pmatrix} 1 & -0.25 \\ -0.25 & 2.2 \end{pmatrix}$$

```
% covariance matrix of the first class
S1 = cov(X1);
```

Fisher Linear Discriminant

- Covariance matrix of the second class:

$$S_2 = \sum_{x\in\omega_2}(x-\mu_2)(x-\mu_2)^T = \left[\begin{pmatrix}9\\10\end{pmatrix}-\begin{pmatrix}8.4\\7.6\end{pmatrix}\right]^2 + \left[\begin{pmatrix}6\\8\end{pmatrix}-\begin{pmatrix}8.4\\7.6\end{pmatrix}\right]^2$$

$$+\left[\begin{pmatrix}9\\5\end{pmatrix}-\begin{pmatrix}8.4\\7.6\end{pmatrix}\right]^2 + \left[\begin{pmatrix}8\\7\end{pmatrix}-\begin{pmatrix}8.4\\7.6\end{pmatrix}\right]^2 + \left[\begin{pmatrix}10\\8\end{pmatrix}-\begin{pmatrix}8.4\\7.6\end{pmatrix}\right]^2$$

$$= \begin{pmatrix}2.3 & -0.05\\ -0.05 & 3.3\end{pmatrix}$$

```
% covariance matrix of the first class
S2 = cov(X2);
```

Fisher Linear Discriminant

- Within-class scatter matrix:

$$S_w = S_1 + S_2 = \begin{pmatrix} 1 & -0.25 \\ -0.25 & 2.2 \end{pmatrix} + \begin{pmatrix} 2.3 & -0.05 \\ -0.05 & 3.3 \end{pmatrix}$$

$$= \begin{pmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{pmatrix}$$

```
% within-class scatter matrix
Sw = S1 + S2 ;
```

Fisher Linear Discriminant

- The LDA projection is then obtained as the solution of the generalized eigen value problem $S_W^{-1} S_B w = \lambda w$

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

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

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

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

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

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

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

Fisher Linear Discriminant

```
% computing the LDA projection
invSw = inv(Sw);

invSw_by_SB = invSw * SB;

% getting the projection vector
[V,D] = eig(invSw_by_SB)

% the projection vector
W = V(:,1);
```

- Hence

$$\begin{pmatrix} 9.2213 & 6.489 \\ 4.2339 & 2.9794 \end{pmatrix} w_1 = \underbrace{0}_{\lambda_1} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

and

$$\begin{pmatrix} 9.2213 & 6.489 \\ 4.2339 & 2.9794 \end{pmatrix} w_2 = \boxed{\underbrace{12.2007}_{\lambda_2}} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

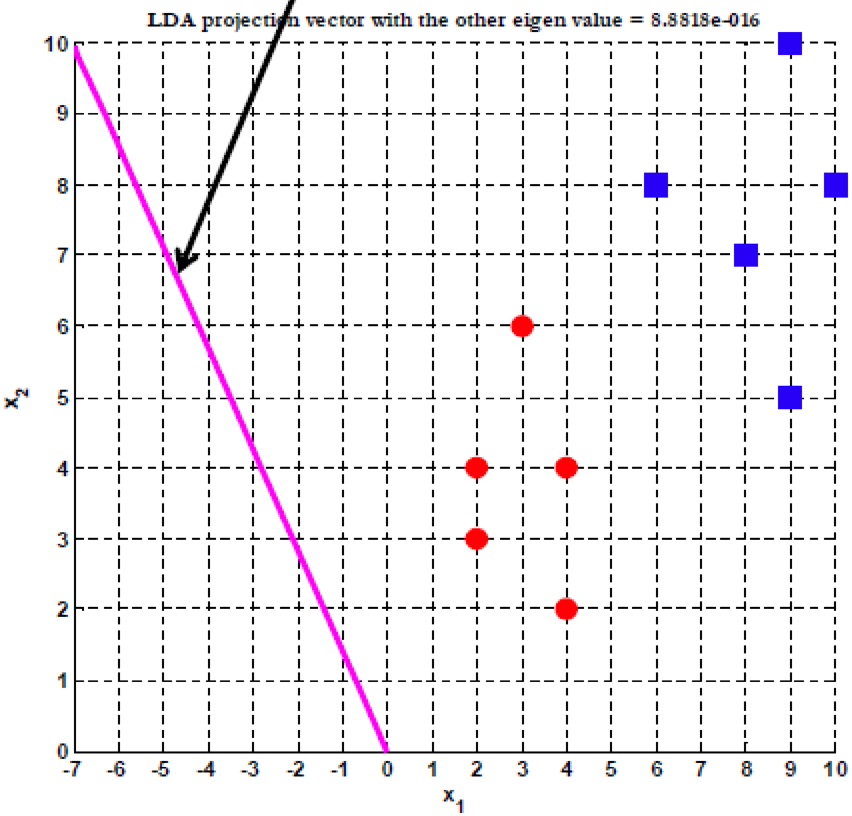
Thus;

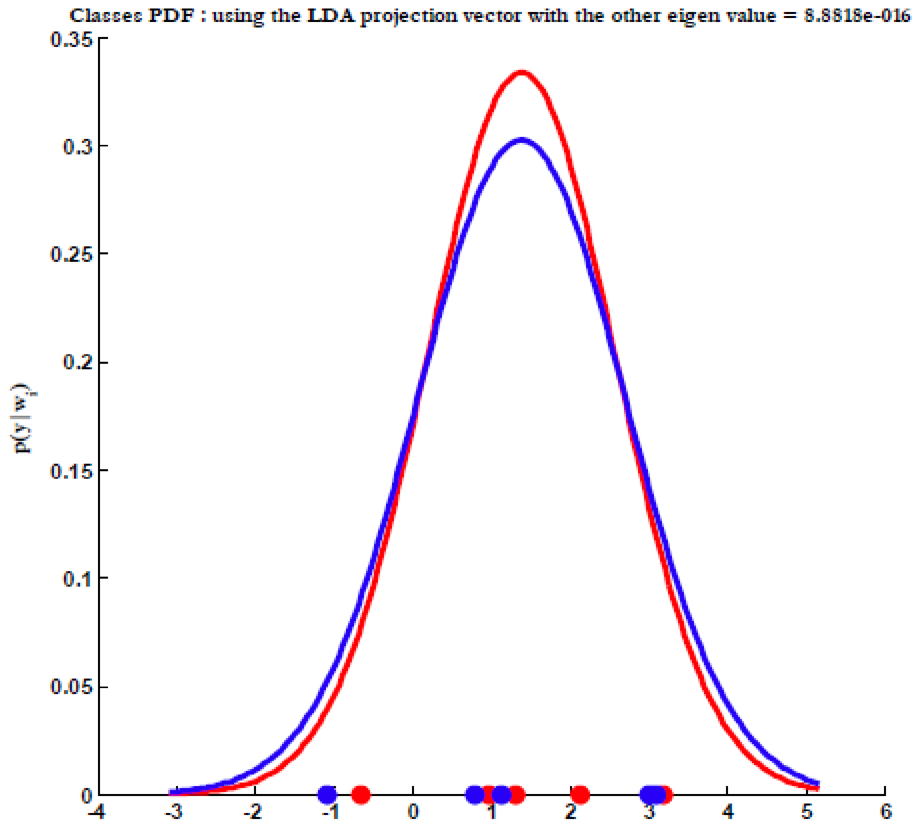
$$w_1 = \begin{pmatrix} -0.5755 \\ 0.8178 \end{pmatrix} \quad and \quad \boxed{w_2 = \begin{pmatrix} 0.9088 \\ 0.4173 \end{pmatrix} = w^*}$$

- The optimal projection is the one that given maximum $\lambda = J(w)$

Fisher Linear Discriminant

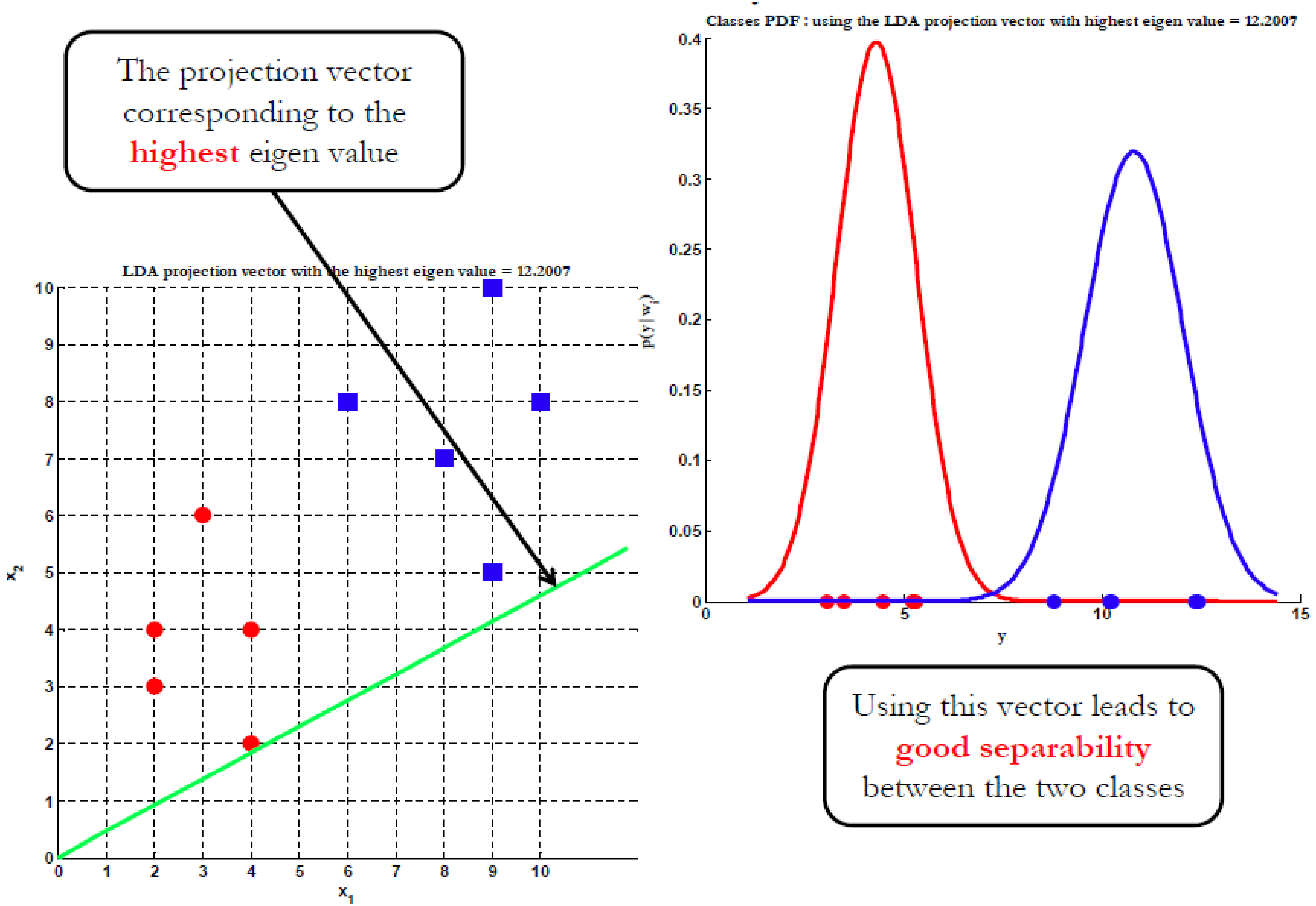
The projection vector corresponding to the **smallest** eigen value





Using this vector leads to **bad separability** between the two classes

Fisher Linear Discriminant



End
Week 16