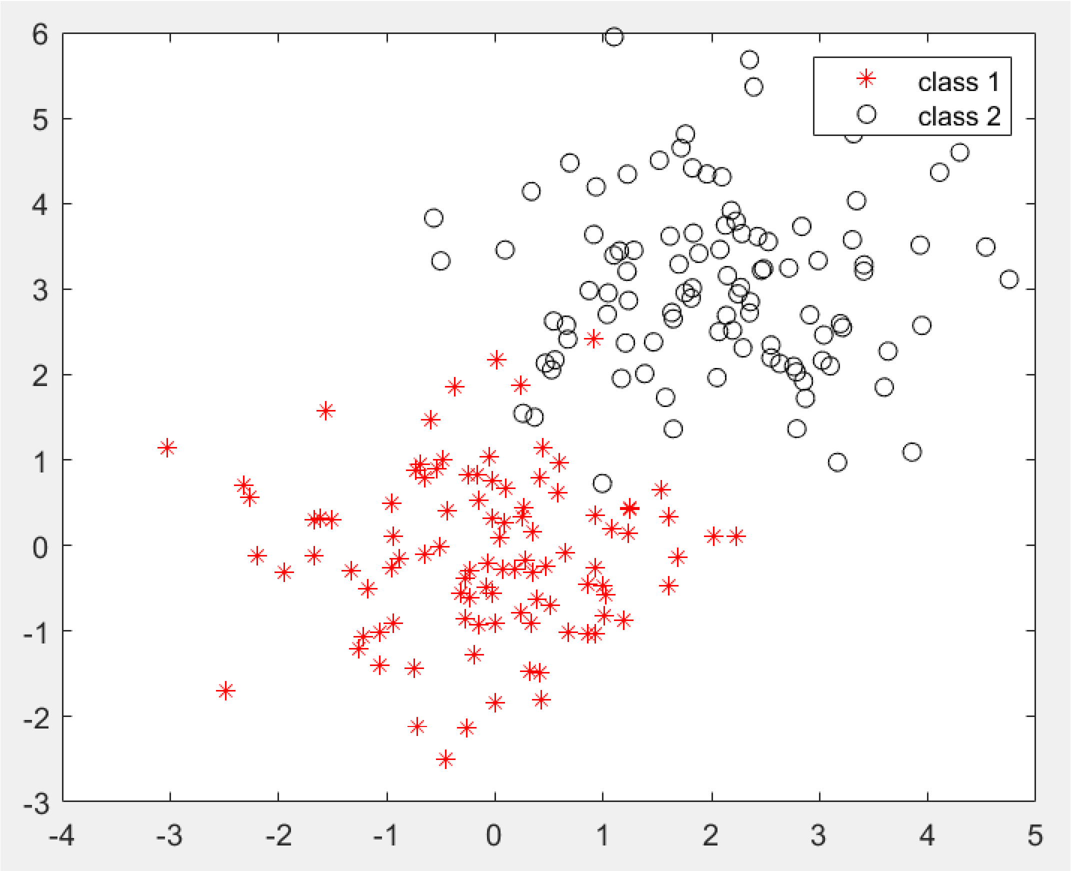
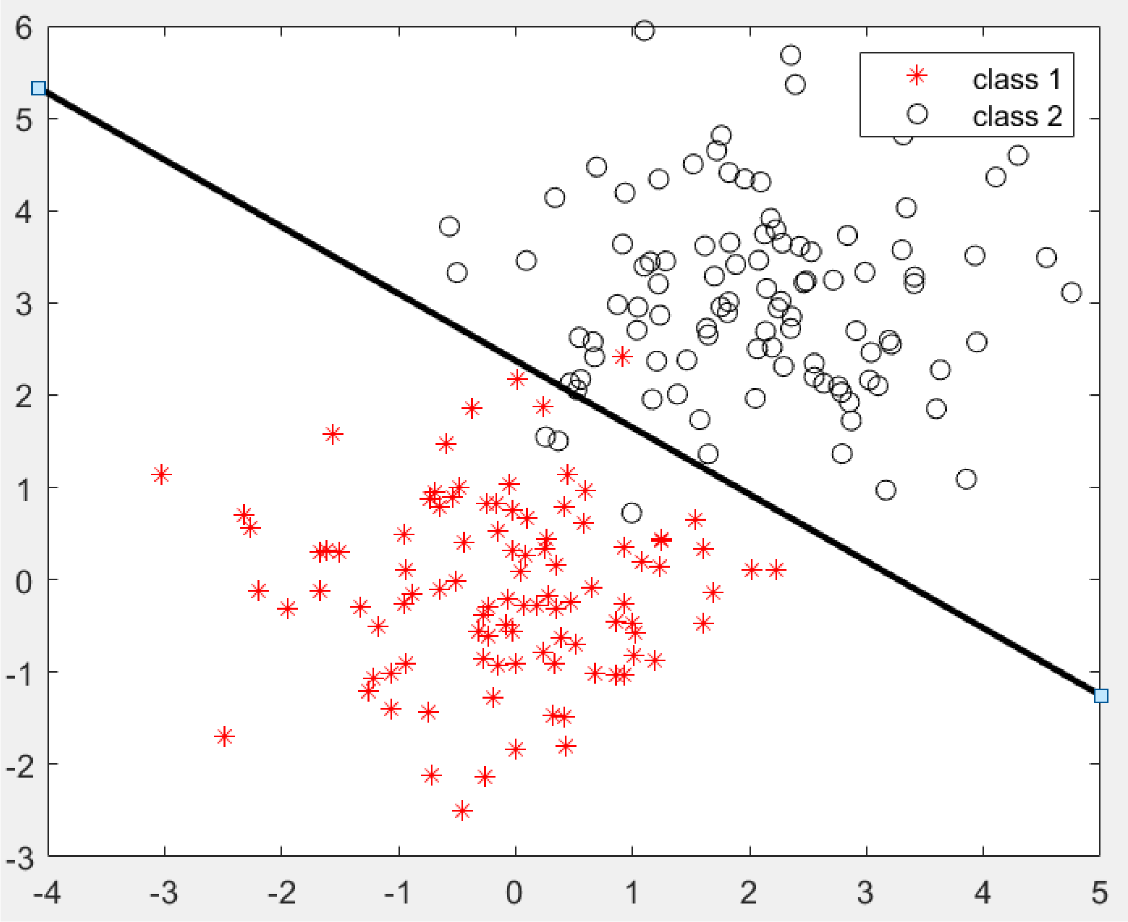
# Linear Discriminant Analysis (LDA)

Consider the following two-class classification problem:



Our objective is to design a linear discriminant function (a line or hyperplane) to separate the samples in the two classes.



The linear discriminant function can be expressed as:

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0$$

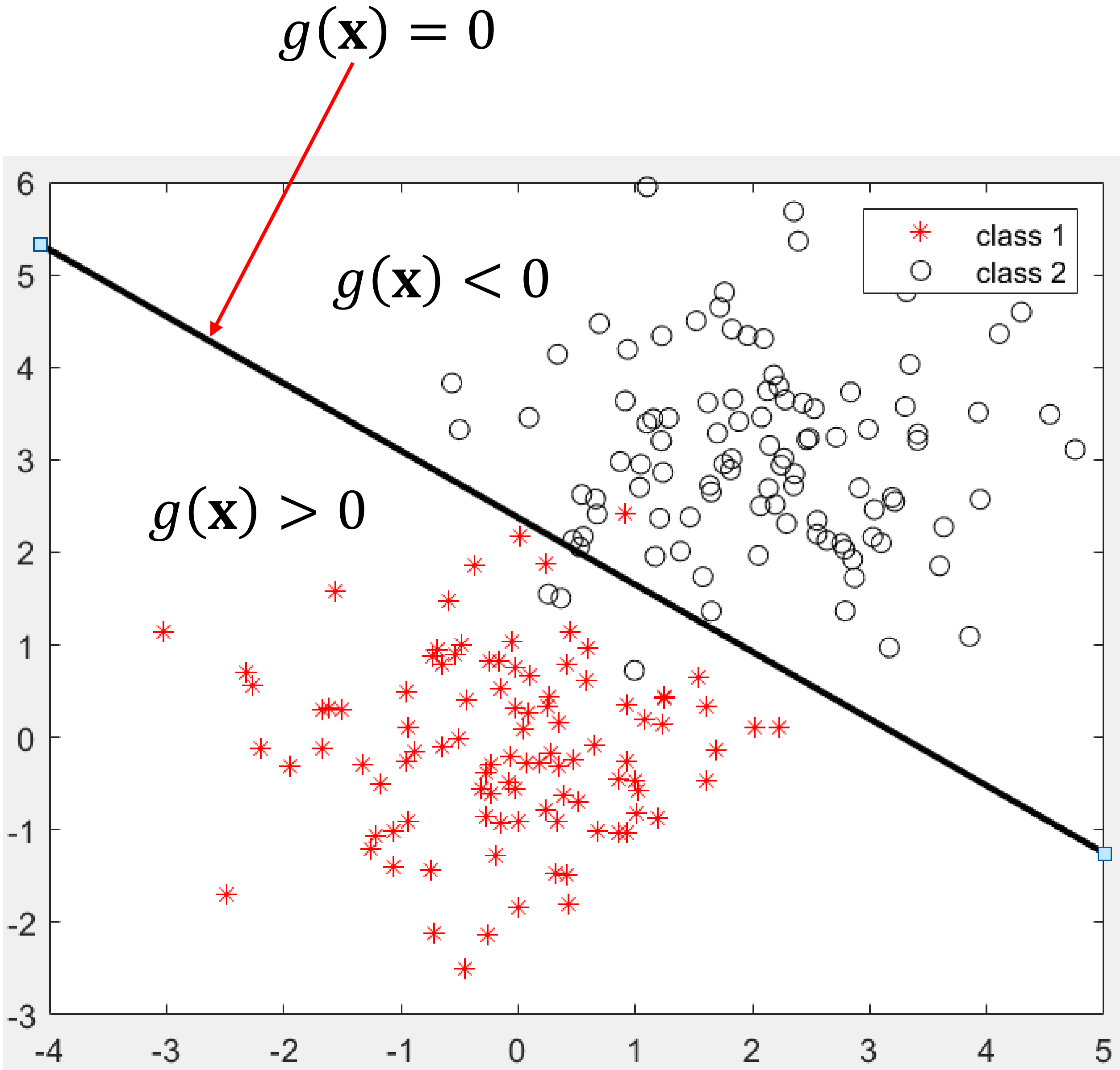
Where $\mathbf{w}$ is the weight vector and $w_0$ is the bias or threshold weight.

For a two-class problem, the classifier implements the following decision rule:

Decide $\omega_1$ if $g(\mathbf{x}) > 0$

Decide $\omega_2$ if $g(\mathbf{x}) < 0$

On decision boundary if $g(\mathbf{x}) = 0$


$g(\mathbf{x}) = 0$
$g(\mathbf{x}) < 0$
$g(\mathbf{x}) > 0$
class 1
class 2
6
5
4
3
2
1
0
-1
-2
-3
-4
-3
-2
-1
0
1
2
3
4
5

Assume $\mathbf{x_1}$ and $\mathbf{x_2}$ are two points on the decision surface, then we have:

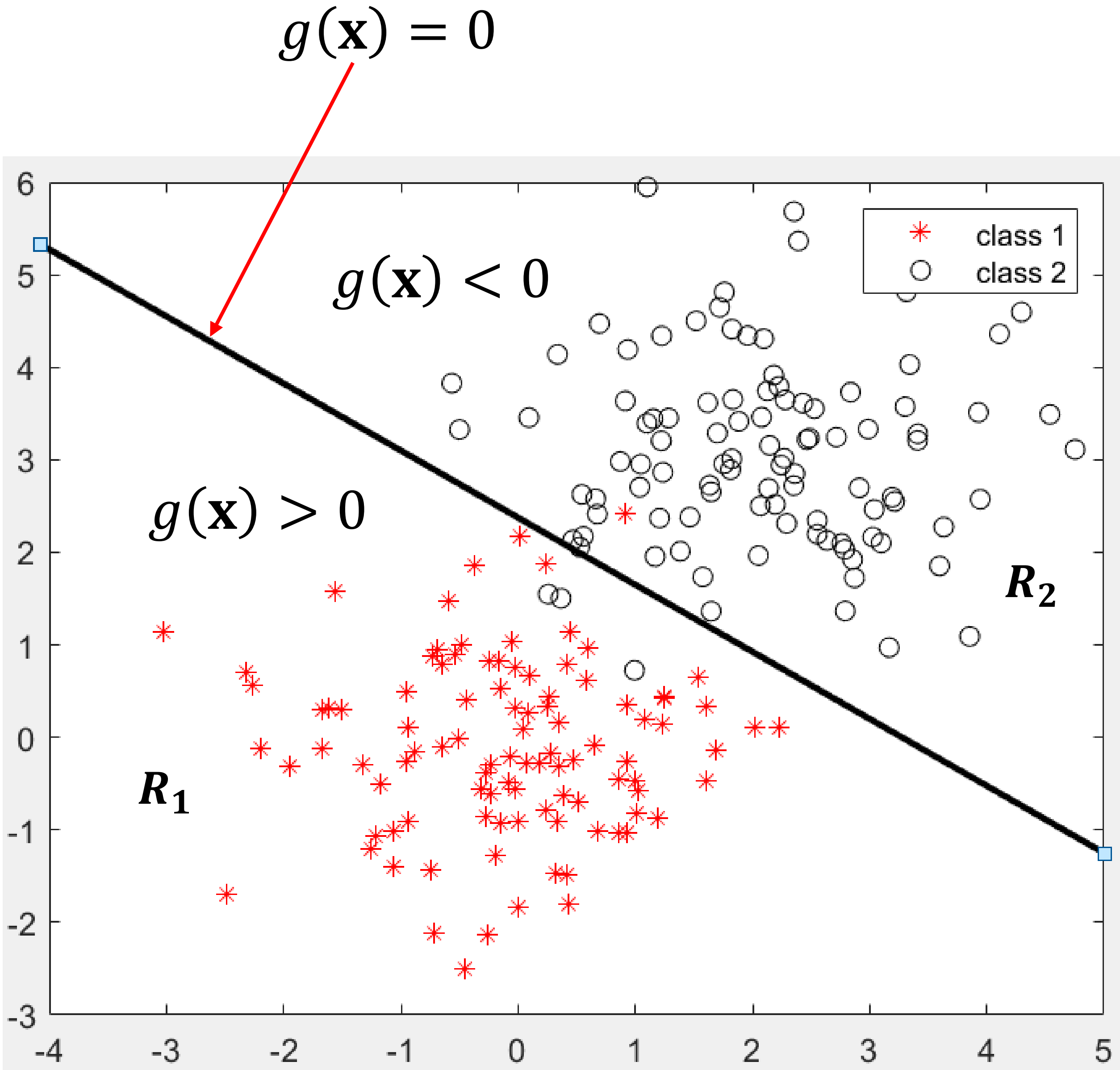
$$\mathbf{w}^{\boldsymbol{T}}\mathbf{x_1} + w_0 = \mathbf{w}^{\boldsymbol{T}}\mathrm{x}_{\mathbf{2}} + w_0$$

or

$$\mathbf{w}^T(\mathbf{x_1} - \mathbf{x_2}) = 0$$

Which means that $\mathbf{w}$ is normal to any vector lying on the hyperplane. In general, the hyperplane $H$ divides the feature space into two half-spaces: Decision region $R_1$ for $\omega_1$ and region $R_2$ for $\omega_2$.

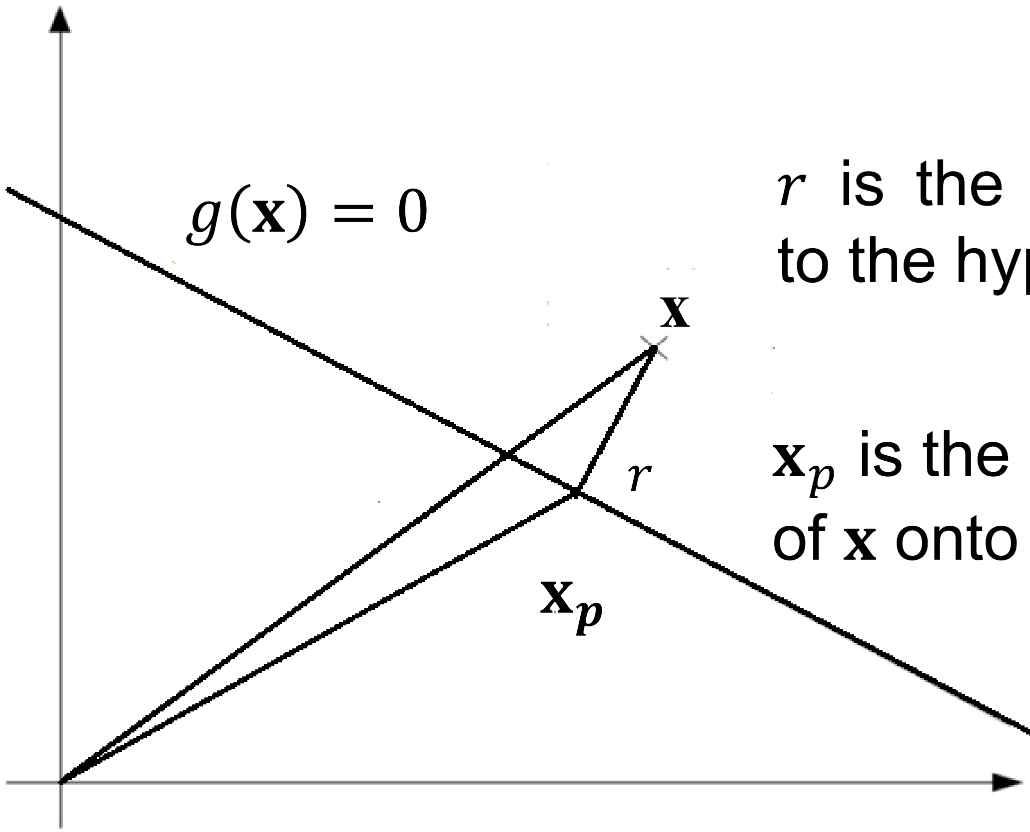
Because $g(\mathbf{x}) > 0$ if $\mathbf{x}$ is in region $R_1$, it follows that the normal vector $\mathbf{w}$ points to $R_1$ and it is said on the positive side of $H$, and any $\mathbf{x}$ in region $R_2$ on the negative side.



$g(\mathbf{x}) = 0$
$g(\mathbf{x}) < 0$
$g(\mathbf{x}) > 0$
class 1
class 2
$R_2$
$R_1$
6
5
4
3
2
1
0
-1
-2
-3
-4
-3
-2
-1
0
1
2
3
4
5

The discriminant function $g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0$ gives a measure of the distance from $\mathbf{x}$ to the hyperplane:

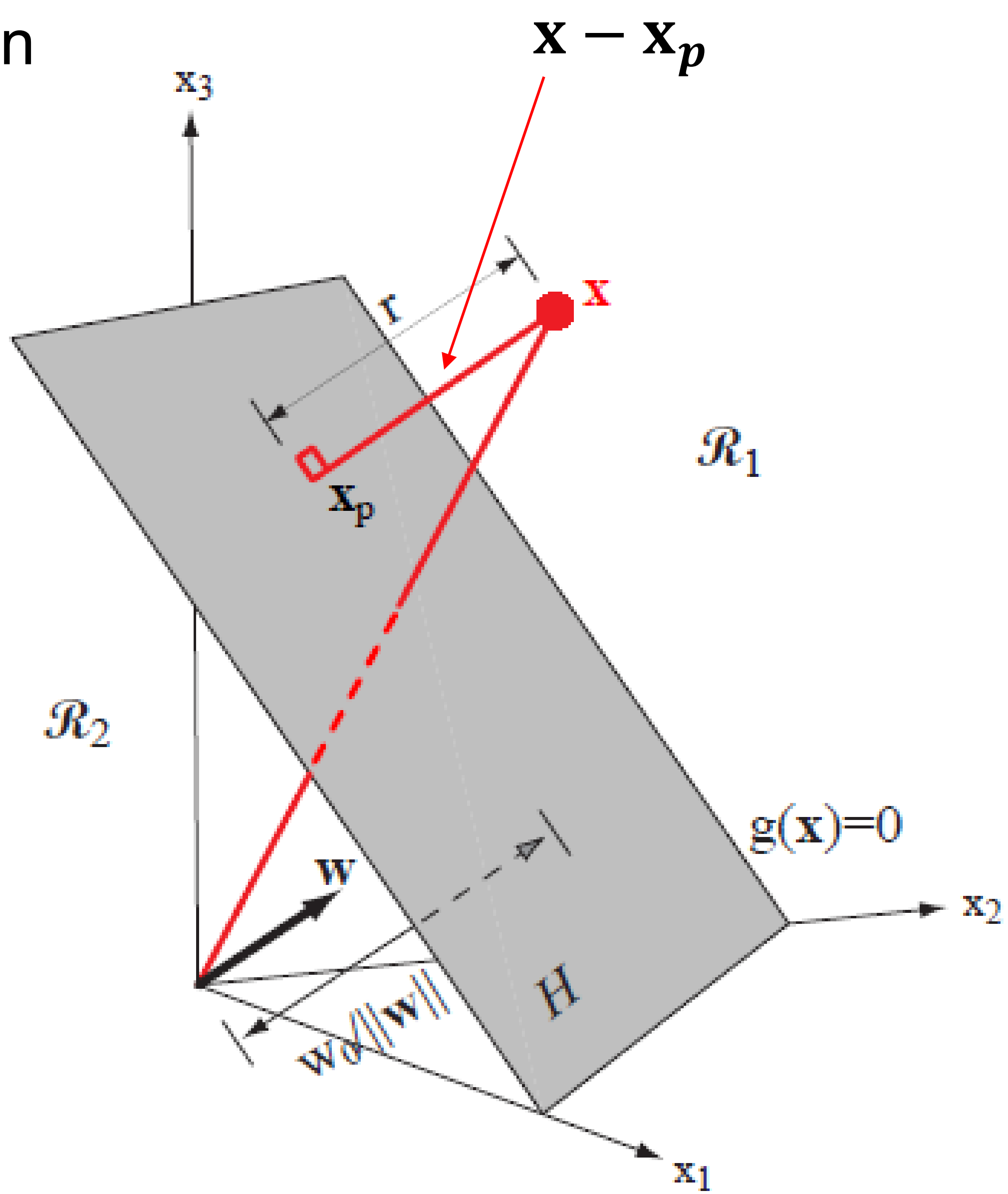
$$\mathbf{x} = \mathbf{x}_p + (\mathbf{x} - \mathbf{x}_p) = \mathbf{x}_p + r\frac{\mathbf{w}}{||\mathbf{w}||}$$




$r$ is the distance from $\mathbf{x}$ to the hyperplane

$\mathbf{x}_p$ is the normal projection of $\mathbf{x}$ onto the hyperplane

# 3D illustration



$r$ is the distance from **x** to the hyperplane. It is

- positive if $\mathbf{x}$ is on the positive side of $H$
- negative if $\mathbf{x}$ is on the negative side of $H$.

$$
\begin{aligned}
g(\mathbf{x}) &= \mathbf{w}^T\mathbf{x} + w_0 \\
&= \mathbf{w}^T\left(\mathbf{x}_p + r\frac{\mathbf{w}}{\|\mathbf{w}\|}\right) + w_0 \\
&= \mathbf{w}^T\mathbf{x}_p + r\frac{\mathbf{w}^T\mathbf{w}}{\|\mathbf{w}\|} + w_0 \\
&= g(\mathbf{x}_p) + r\|\mathbf{w}\| \\
&= r\|\mathbf{w}\|
\end{aligned}
$$

Thus,

$$
r = \frac{g(\mathbf{x})}{\|\mathbf{w}\|}
$$

The distance from the origin (i.e. $\mathbf{x} = 0$) to $H$ is given by:

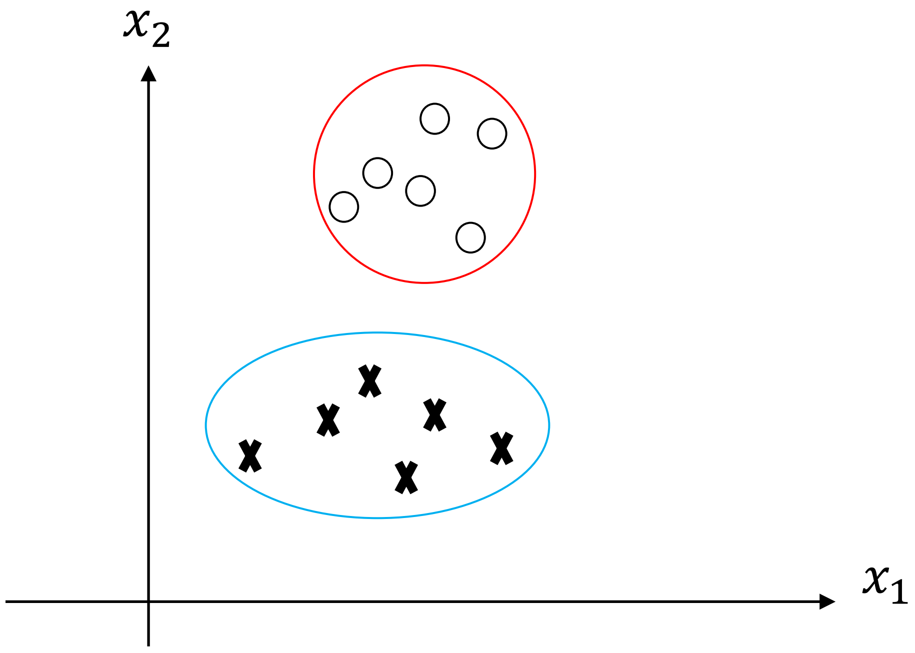
$$r = \frac{g(\mathbf{x})}{||\mathbf{w}||} = \frac{g(\mathbf{0})}{||\mathbf{w}||} = \frac{w_0}{||\mathbf{w}||}$$

Therefore $w_0$ determines the position of the hyperplane $H$, while $\mathbf{w}$ determines the orientation of the hyperplane $H$.

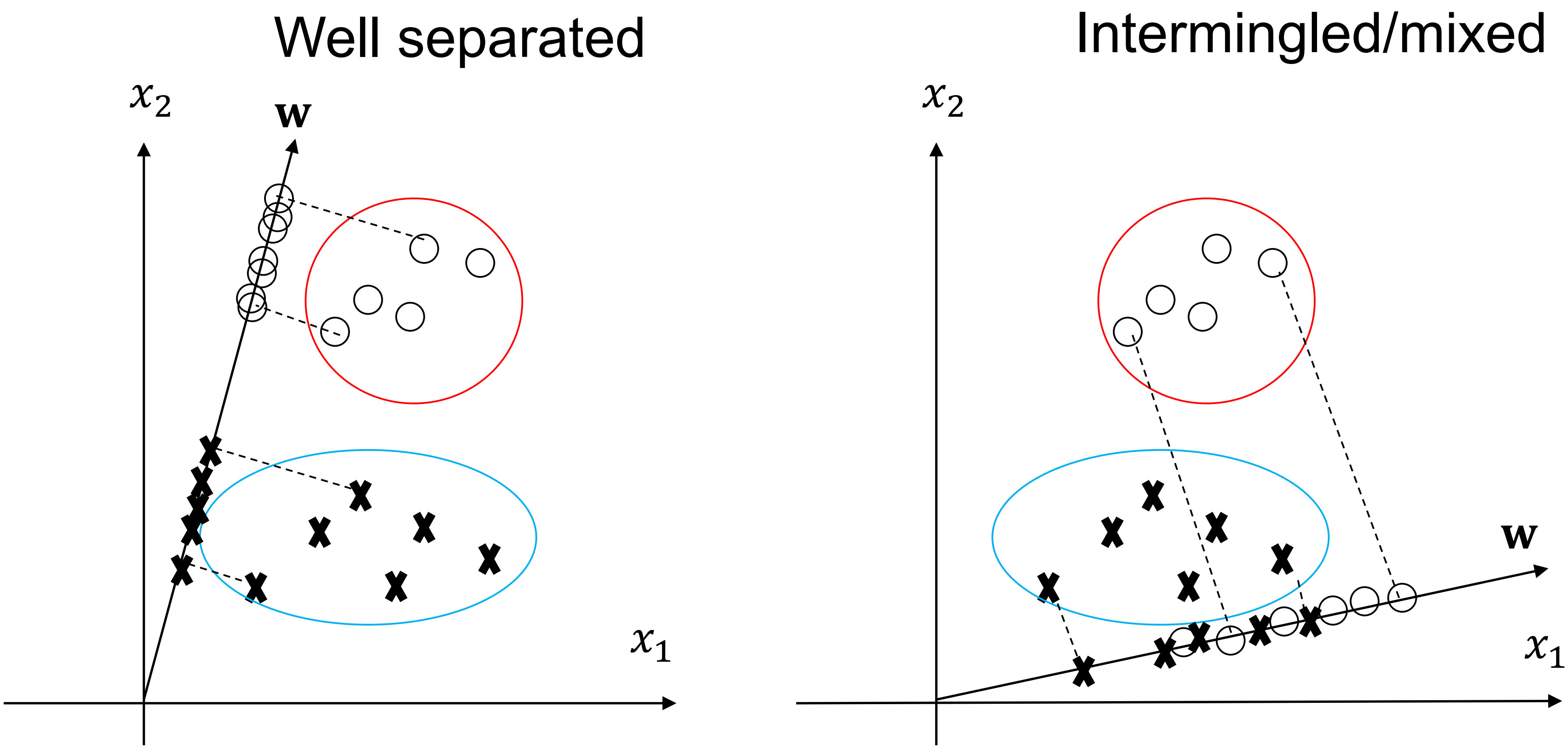
1) If $w_0 > 0$, the origin is on the positive side of $H$
2) If $w_0 < 0$, the origin is on the negative side of $H$
3) If $w_0 = 0$, the hyperplane $H$ passes through the origin.

Geometrically, $\mathbf{w}^T\mathbf{x}$ is the projection of sample $\mathbf{x}$ on the hyperplane in the direction of $\mathbf{w}$. Actually, the magnitude of $\mathbf{w}$ is of no real significance because it merely scale the projection, but the direction of $\mathbf{w}$ is important.

Assume samples in one class fall more or less into one cluster while those in another class fall in another cluster:



We hope the projections of samples on the hyperplane to be well separated as shown in the left figure below, rather than to be intermingled as shown in the right figure:



We next find the $\mathbf{w}$ that will enable accurate classification. A measure of separation between the projected points of two classes is the difference of the sample means.

Assume $\mathbf{m}_i$ is the $d$-dimensional sample mean of class $i$ given by:

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{x}$$

Then the sample mean for the corresponding projected points is given by:

$$\widetilde{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{w}^T \mathbf{x} = \mathbf{w}^T \mathbf{m}_i$$

The distance between the projected means is given by:

$$|\widetilde{m}_1 - \widetilde{m}_2| = \left|\mathbf{w}^{\mathbf{T}}(\mathbf{m_1} - \mathbf{m_2})\right|$$

To obtain good separation of the projected data, we want the difference between means to be large relative to some measures of the spread or scatter of data in each class, which is defined as:

$$\tilde{s}_i^2 = \sum_{\mathbf{x} \in D_i} [g(\mathbf{x}) - \widetilde{m}_i]^2$$

We define the totally within-class scatter of the projected samples as:

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

The Fisher linear discriminant attempts to find such $\mathbf{w}$ that the following criterion is maximized:

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

To derive $J(\mathbf{w})$ as an explicit function of $\mathbf{w}$, we define the scatter matrices as:

$$\mathbf{S}_i = \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T$$

and

$$\boldsymbol{S}_W = \boldsymbol{S}_1 + \boldsymbol{S}_2$$

Then we have:

$$\tilde{s}_i^2 = \sum_{\mathbf{x}\in D_i} (\mathbf{w}^T\mathbf{x} - \mathbf{w}^T\mathbf{m}_i)^2$$

$$= \sum_{\mathbf{x}\in D_i} \mathbf{w}^T(\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T\mathbf{w}$$

$$= \mathbf{w}^T\ \mathbf{S}_i\mathbf{w}$$

Therefore the total within-class scatter can be expressed as :

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

$$= \mathbf{w}^T\ (\mathbf{S}_1 + \mathbf{S}_2)\mathbf{w}$$

$$= \mathbf{w}^T\ \mathbf{S}_W\mathbf{w}$$

Similarly, the separation of the projected sample means can be expressed as:

$$
\begin{aligned}
(\widetilde{m}_1 - m_2)^2 &= (\mathbf{w}^T \mathbf{m}_1 - \mathbf{w}^T \mathbf{m}_2)^2 \\
&= \mathbf{w}^T (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{w} \\
&= \mathbf{w}^T \mathbf{S}_B \mathbf{w}
\end{aligned}
$$

where $\boldsymbol{S}_B$ is the between-class scatter matrix:

$$\boldsymbol{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$$

Thus, the maximization criterion $J(\mathbf{w})$ can be expressed as:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}$$

This expression is known as the generalized Rayleigh quotient, and the vector that maximizes the quotient must satisfy:

$$\mathbf{S}_B\mathbf{w} = \lambda\mathbf{S}_W\mathbf{w}$$

for some constant $\lambda$, which is a generalized eigenvalue problem.

If $\mathbf{S}_W$ is non-singular, we can obtain a conventional eigenvalue problem:

$$\mathbf{S}_W^{-1}\mathbf{S}_B\mathbf{w} = \lambda\mathbf{w}$$

By solving the eigenvalue problem, we can get the eigen vector $\mathbf{w}$.

- **Eigenvalue problem vs generalized eigenvalue problem**

  - ❑ Eigenvalue problem

$$\mathbf{Ax} = \lambda\mathbf{x}$$

  - ❑ Generalized eigenvalue problem

$$\mathbf{Ax} = \lambda\mathbf{Bx}$$

For this particular problem, actually it is unnecessary to solve for the eigenvalues and eigenvectors of $\mathbf{S}_W^{-1}\mathbf{S}_B$.

$$\boldsymbol{S}_B\mathbf{w} = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T\mathbf{w} = (\mathbf{m}_1 - \mathbf{m}_2)\rho$$

where $\rho$ is the difference of sample mean projections on the hyperplane:

$$\rho = (\mathbf{m}_1 - \mathbf{m}_2)^T\mathbf{w} = \mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)$$

Substitute $\boldsymbol{S}_B\mathbf{w}$ into the following equation:

$$\mathbf{S}_B\mathbf{w} = \lambda\mathbf{S}_W\mathbf{w}$$

we obtain:

$$(\mathbf{m}_1 - \mathbf{m}_2)\rho \; = \lambda\mathbf{S}_W\mathbf{w}$$

$\rho$ and $\lambda$ can be considered as scaling factors. Since we are only concerned with the direction of $\mathbf{w}$, we can ignore the two scaling factor to yield:

$$\mathbf{w} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$$

If matrix $\boldsymbol{S}_W$ is singular, i.e.

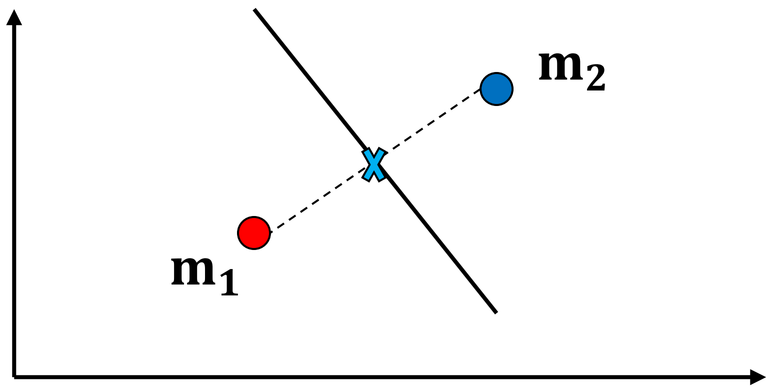
$$|\boldsymbol{S}_W| \approx 0$$

Then we can use the so-called regularization technique:

$$\mathbf{w} = (\mathbf{S}_W + \beta\mathbf{I})^{-1}\,(\mathbf{m}_1 - \mathbf{m}_2)$$

Where $\beta$ is a small positive number, say 0.001, and $\mathbf{I}$ is an identity matrix.

The bias or threshold $w_0$ is often so defined that the middle of two class means is on the hyperplane:




$$\mathbf{w}^T\mathbf{x} + w_0 = \mathbf{w}^{\boldsymbol{T}}\frac{\mathbf{m}_1 + \mathbf{m}_2}{\mathbf{2}} + w_0 = 0$$

$$w_0 = -\frac{\mathbf{w}^T(\mathbf{m}_1 + \mathbf{m}_2)}{2}$$

**Note:**

The above way of finding the bias term $w_0$ is based on the assumption that the sample projections on $\mathbf{w}$ of the two classes have the **same scatter or spread**.

If this condition is not met, we may **adjust $\boldsymbol{w_0}$** to get better performance.

## Summary of linear discriminant analysis

1) Calculate the mean vectors $\mathbf{m}_1$ and $\mathbf{m}_2$ of samples in the two classes, respectively:

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{x}$$

2) Calculate the within-class scatter matrix $\boldsymbol{S}_W$

$$\mathbf{S}_i = \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T$$

$$\boldsymbol{S}_W = \boldsymbol{S}_{\mathbf{1}} + \boldsymbol{S}_{\mathbf{2}}$$

3) Calculate $\mathbf{w}$ and $w_0$

$$\mathbf{w} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$$

$$w_0 = -\frac{\mathbf{w}^T(\mathbf{m}_1 + \mathbf{m}_2)}{2}$$

4) Construct discriminant function and classify samples

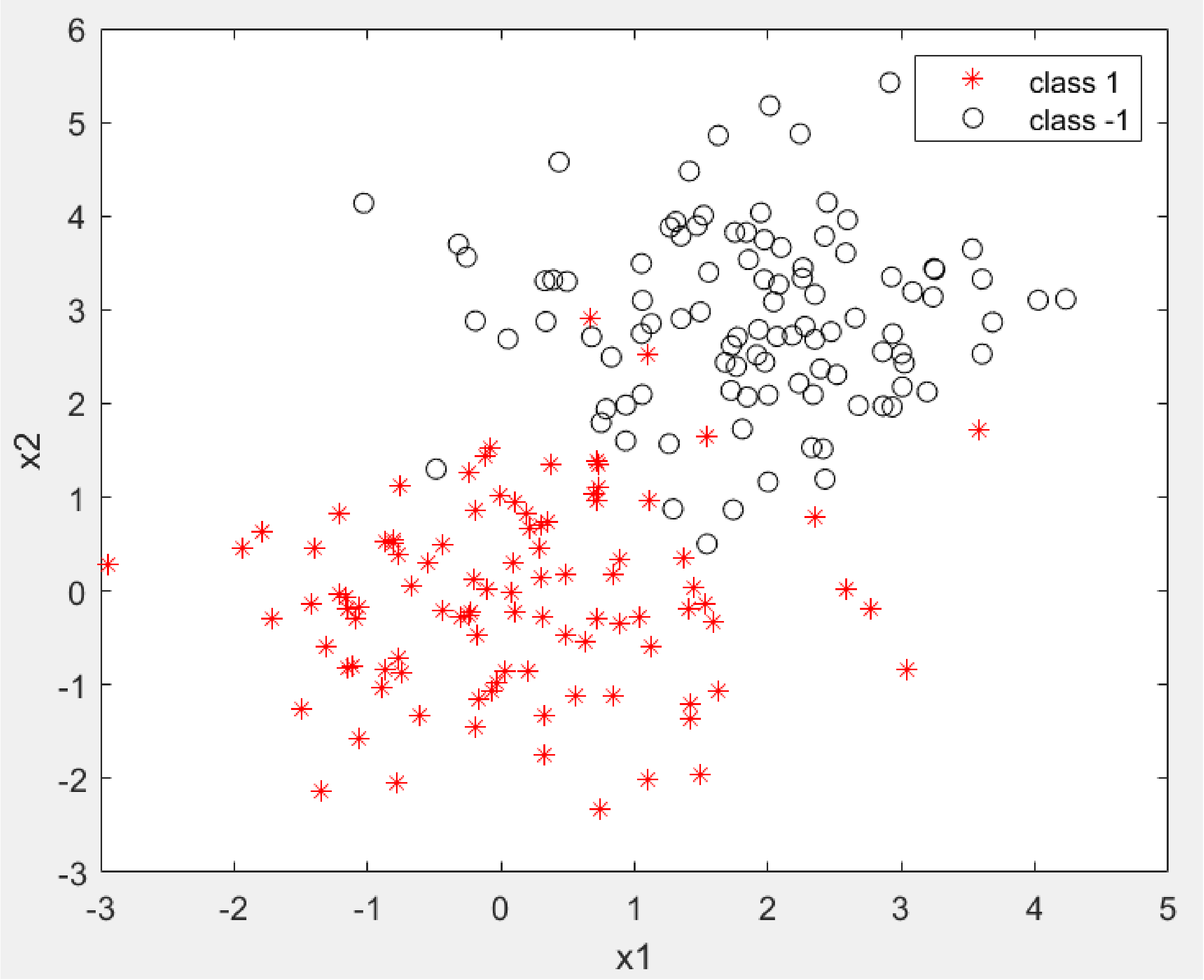
$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0$$

Decide $\omega_1$ if $g(\mathbf{x}) > 0$

decide $\omega_2$ if $g(\mathbf{x}) < 0$

On decision boundary if $g(\mathbf{x}) = 0$

Consider the following samples in two classes:



$$\mathbf{m}_1 = \begin{bmatrix} 0.1083 \\ -0.0653 \end{bmatrix} \qquad \mathbf{m}_2 = \begin{bmatrix} 1.8945 \\ 2.9026 \end{bmatrix}$$

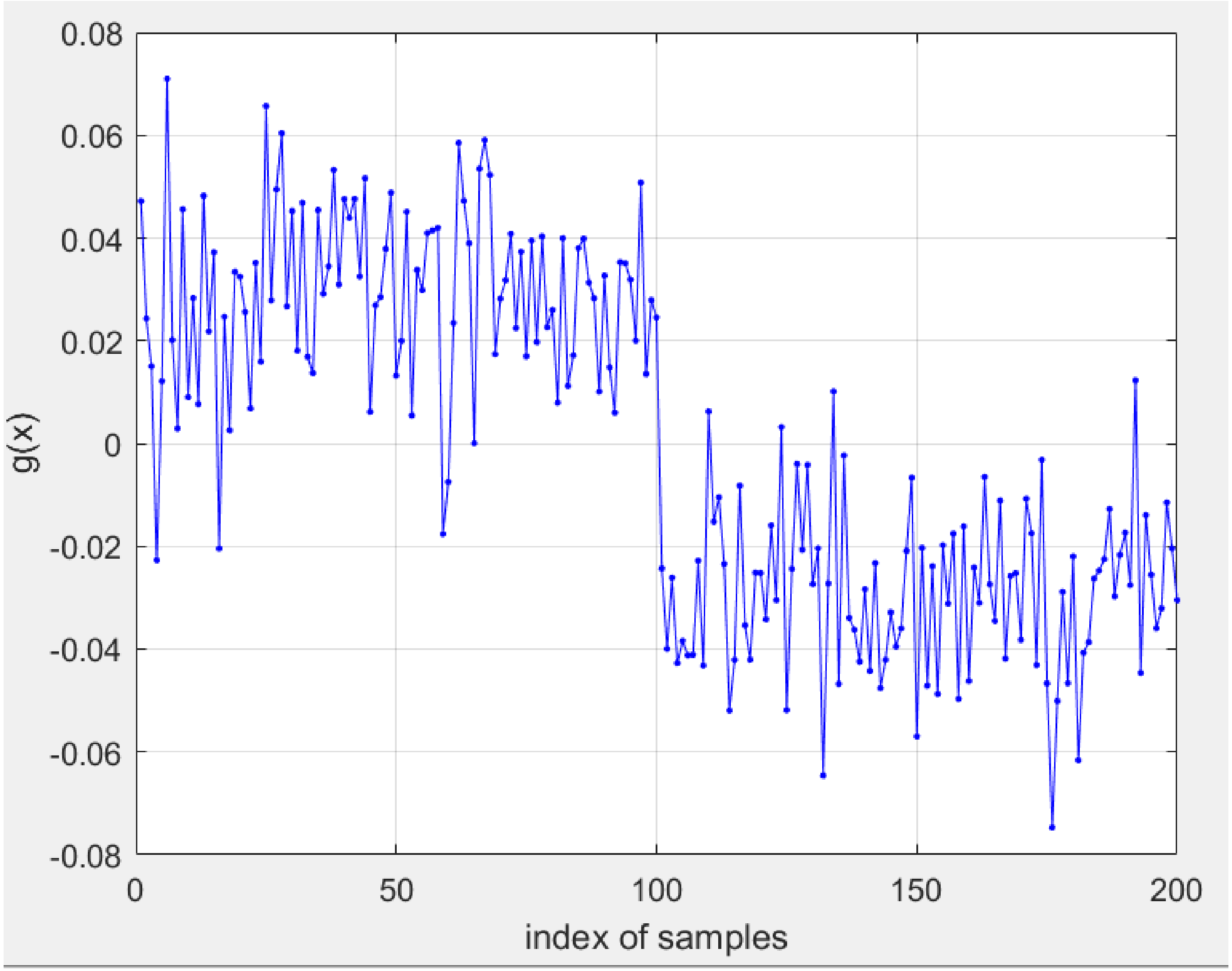
$$\mathbf{S}_W = \begin{bmatrix} 228.9365 & 10.9883 \\ 10.9883 & 189.216 \end{bmatrix}$$

$$\mathbf{w} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) = \begin{bmatrix} -0.0071 \\ -0.0153 \end{bmatrix}$$

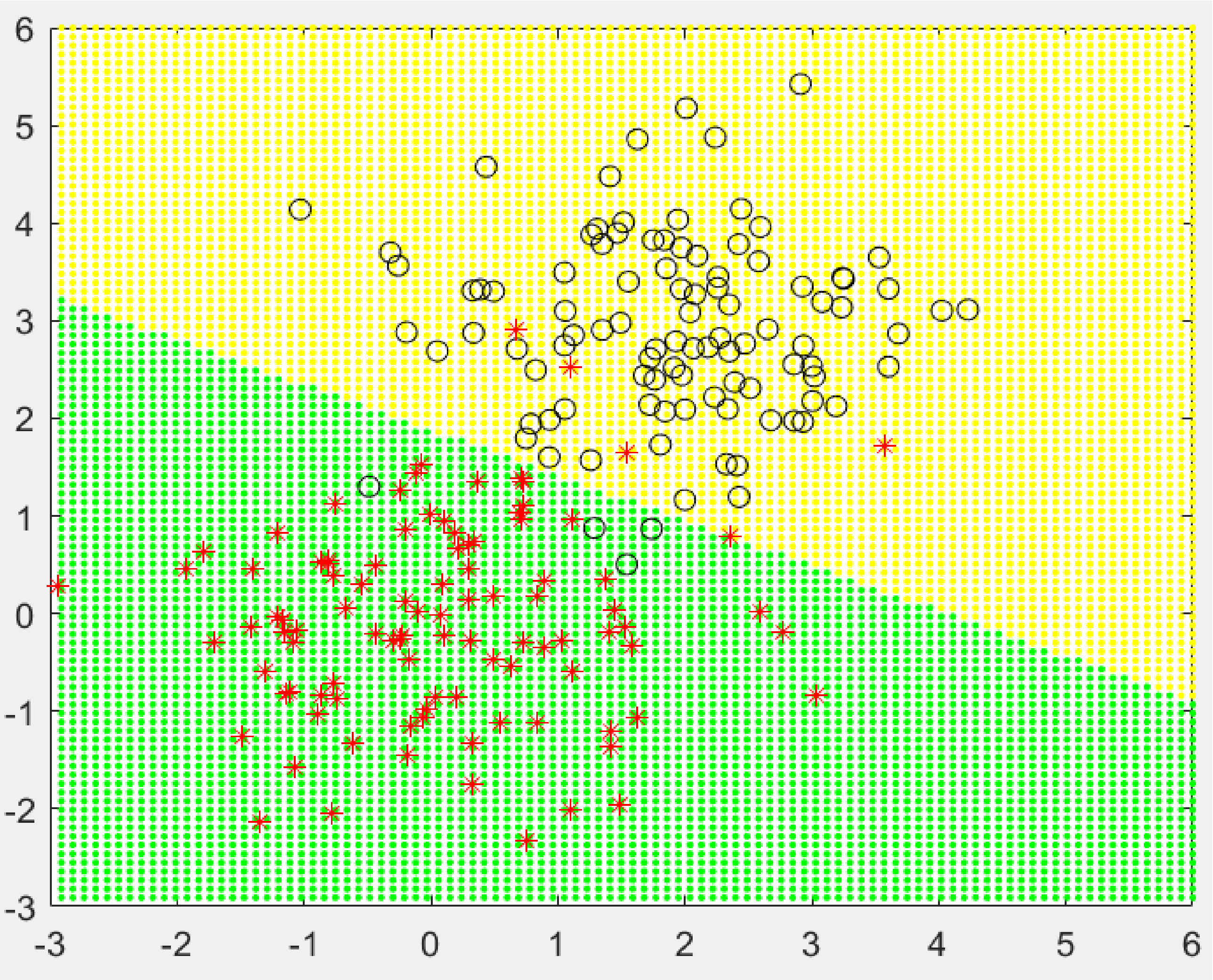
$$w_0 = 0.0288$$

$$\begin{aligned} g(\mathbf{x}) &= [-0.0071 \quad -0.0153]\mathbf{x} + 0.0288 \\ &= -0.0071x_1 - 0.0153x_2 + 0.0288 \end{aligned}$$

# Discriminant function g(**x**)



# Decision boundary



We can also find $\mathbf{w}$ by solving the following eigenvalue problem:

$$\mathbf{S}_W^{-1}\mathbf{S}_B\mathbf{w} = \lambda\mathbf{w}$$

The eigenvalue problem can be solved using the Matlab function "eig":

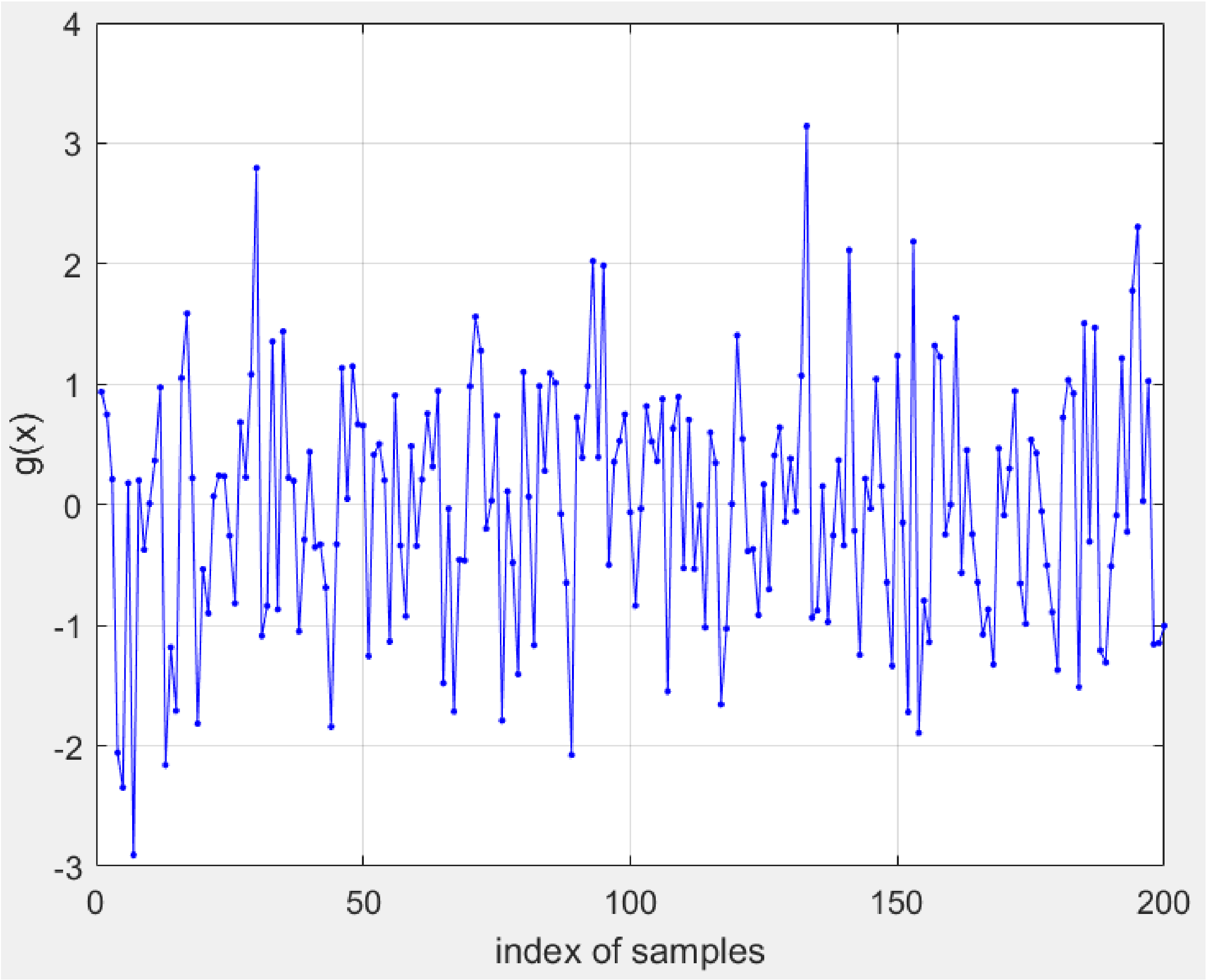
$$[\mathbf{v}, \mathbf{d}] = eig(inv(\mathbf{S}_W) * \mathbf{S}_B)$$

$$\mathbf{d} = \begin{bmatrix} 0.0000 & 0 \\ 0 & 0.0580 \end{bmatrix}$$

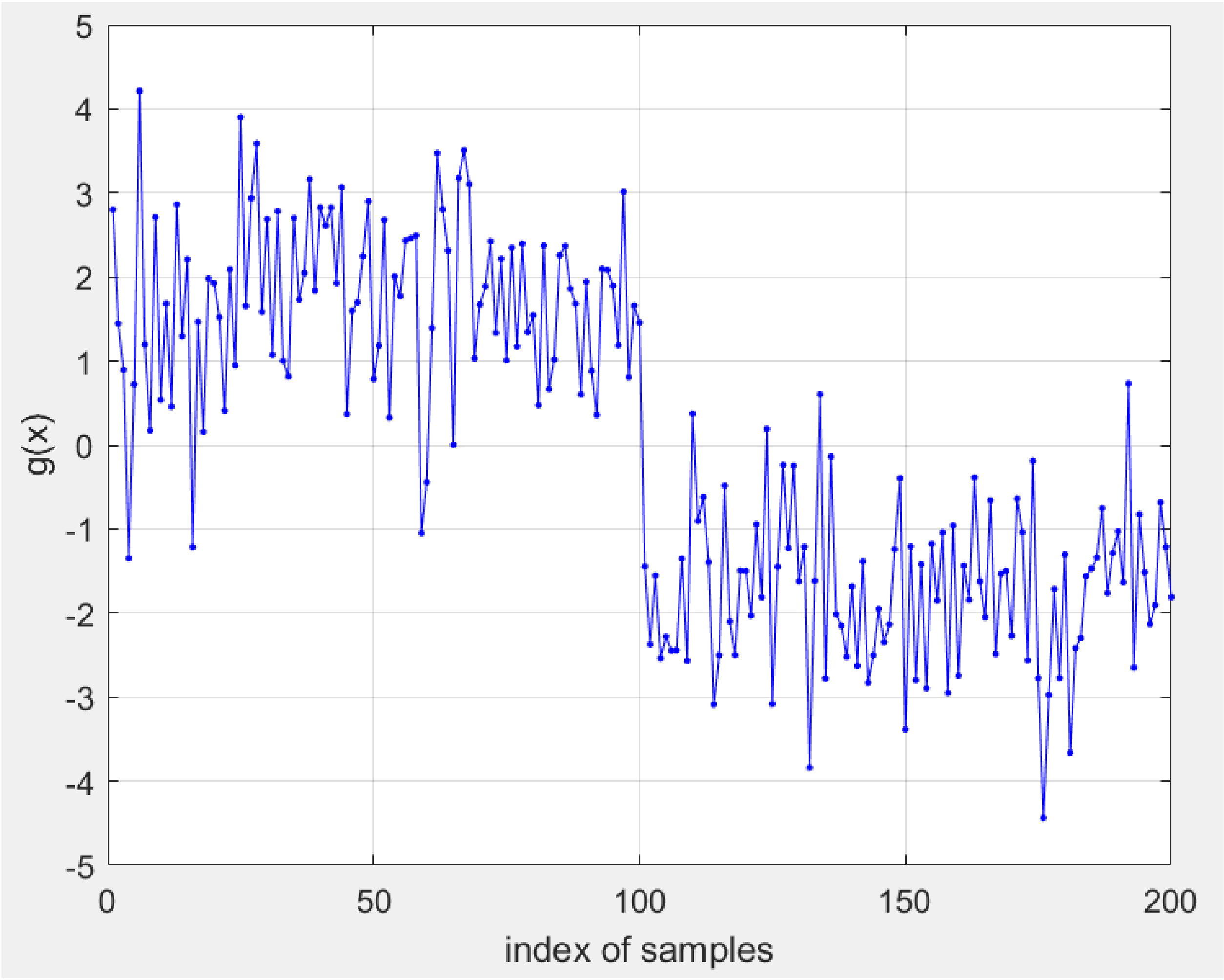
$$\mathbf{v} = \begin{bmatrix} -0.8568 & -0.4198 \\ 0.5157 & -0.9076 \end{bmatrix}$$

Elements on the main diagonal of $\mathbf{d}$ are the eigenvalues and column vectors of $\mathbf{v}$ are the eigenvectors.

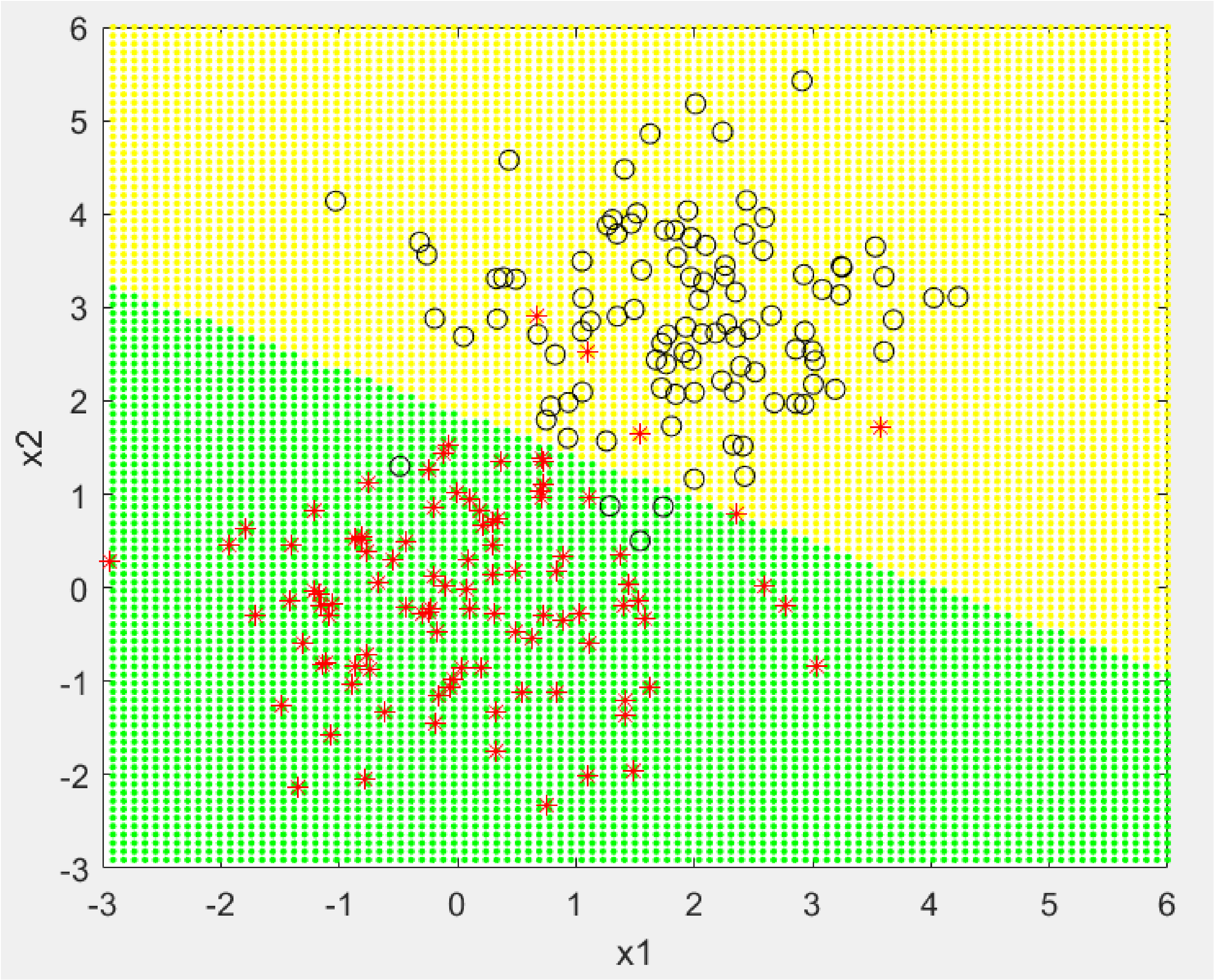
The first eigenvector corresponding to eigenvalue of 0.000 has no discriminative power. If we use it as $\mathbf{w}$, $g(\mathbf{x})$ is as follow:



If the second eigenvector corresponding to eigenvalue of 0.058 is used as $\mathbf{w}$, the $g(\mathbf{x})$ is as follow:



# Decision boundary



**Notes:**

(1) The $\mathbf{w}$ found by the eigenvalue method does not guarantee that samples in class 1 are on the positive side of the decision boundary. Samples of class one with class label 1 falls on the negative side, we replace $\mathbf{w}$ and $w_0$ with $-\mathbf{w}$ and -$w_0$ respectively.

(2) The $\mathbf{w}$ found by the direct calculation method and the (generalized) eigenvalue method usually have different values, but their directions are the same or similar.

For the example above, by using the direct calculation method, we obtained:

$$\mathbf{w} = \begin{bmatrix} -0.0071 \\ -0.0153 \end{bmatrix} \qquad \frac{\mathbf{w}(1)}{\mathbf{w}(2)} = \frac{-0.0071}{-0.0153} = 0.4625$$

By using the eigenvalue method, we obtained:

$$\mathbf{w} = \begin{bmatrix} -0.4198 \\ -0.9076 \end{bmatrix}$$

$$\frac{\mathbf{w}(1)}{\mathbf{w}(2)} = \frac{-0.4198}{-0.9076} = 0.4625$$

Obviously, the directions of the two $\mathbf{w}$ produced by the two methods are the same, though they look very different.

# Multiple Discriminant Analysis

For a $c-$class problem, the natural generalization of Fisher's linear discriminant involves $c$ -1 discriminant functions. The projection is thus from a $d$-dimensional space to a $c-1$ dimensional space. Assuming $d \geq c$.

The generalization for the within-class scatter matrix is as:

$$\boldsymbol{S}_W = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{c}} \boldsymbol{S}_{\boldsymbol{i}}$$

$$\mathbf{S}_i = \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T$$

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{x}$$

We define the total mean vector as:

$$\mathbf{m} = \frac{1}{n}\sum \mathbf{x} = \frac{1}{n}\sum_{i=1}^{c} n_i \mathbf{m}_i$$

and total scatter matrix:

$$\mathbf{S}_T = \sum (\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T$$

$$= \sum_{i=1}^{c} \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i + \mathbf{m}_i - \mathbf{m})(\mathbf{x} - \mathbf{m}_i + \mathbf{m}_i - \mathbf{m})^T$$

$$= \sum_{i=1}^{c} \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T + \sum_{i=1}^{c} \sum_{\mathbf{x} \in D_i} (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T$$

$$= \mathbf{S}_W + \textstyle\sum_{i=1}^{c} n_i \, (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T$$

The second term can be considered as the between-class scatter matrix:

$$\mathbf{S}_B = \sum_{i=1}^{c} n_i \, (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T$$

The projection from a *d*-dimensional space to a $c-1$ dimensional space is by $c-1$ discriminant functions:

$$g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} \qquad\qquad i = 1,2, \dots, c-1$$

Define:

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) \\ \vdots \\ g_{c-1}(\mathbf{x}) \end{bmatrix}$$

$$\mathbf{W}^T = \begin{bmatrix} \mathbf{w}_1^T \\ \vdots \\ \mathbf{w}_{c-1}^T \end{bmatrix}$$

Then the projection from $d$-dimensional space to $c-1$ dimensional space can be expressed as:

$$\mathbf{g}(\mathbf{x}) = \mathbf{W}^T\mathbf{x}$$

Samples $\mathbf{x}_1$, $\mathbf{x}_2, \cdots,$ $\mathbf{x}_n$ are projected to a set of new representations $\mathbf{g}(\mathbf{x}_1)$, $\mathbf{g}(\mathbf{x}_2)$, $\cdots$, $\mathbf{g}(\mathbf{x}_n)$, which can also be described by their mean vectors and scatter matrices. Thus, we define:

$$\widetilde{\mathbf{m}}_i = \frac{1}{n_i}\sum_{\mathbf{x}\in D_i}\mathbf{g}(\mathbf{x})$$

$$\widetilde{\mathbf{m}} = \frac{1}{n}\sum_{i=1}^{c} n_i\widetilde{\mathbf{m}}_i$$

$$\tilde{\mathbf{S}}_W = \sum_{i=1}^{c} \sum_{x \in D_i} (\mathbf{g}(\mathbf{x}_i) - \tilde{\mathbf{m}}_i)(\mathbf{g}(\mathbf{x}_i) - \tilde{\mathbf{m}}_i)^T$$

$$\tilde{\mathbf{S}}_B = \sum_{i=1}^{c} n_i (\tilde{\mathbf{m}}_i - \tilde{\mathbf{m}})(\tilde{\mathbf{m}}_i - \tilde{\mathbf{m}})^T$$

Then we can obtain:

$$\tilde{\mathbf{S}}_W = \mathbf{W}^T \mathbf{S}_W \mathbf{W}$$

$$\tilde{\mathbf{S}}_B = \mathbf{W}^T \mathbf{S}_B \mathbf{W}$$

The objective here is to find such $\mathbf{W}$ that the ratio of the between-class scatter and within-class scatter is maximized. A simple measure of the scatter is the determinant of the scatter matrix. With this measure, the criterion function is given by:

$$J(\mathbf{W}) = \frac{|\tilde{\mathbf{S}}_B|}{|\tilde{\mathbf{S}}_W|} = \frac{|\mathbf{W}^T \mathbf{S}_B \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_W \mathbf{W}|}$$

The problem of finding a rectangular matrix that maximizes $J$ is not easy. Fortunately, it has been approved that the columns of **W** are the generalized eigenvectors corresponding to the largest eigenvalues of $\boldsymbol{S}_B$ and $\boldsymbol{S}_W$:

$$\mathbf{S}_B \mathbf{w}_i = \lambda_i \mathbf{S}_W \mathbf{w}_i$$

which can be solved using software toolbox such as Matlab.

If $\boldsymbol{S}_W$ is non-singular, then we have:

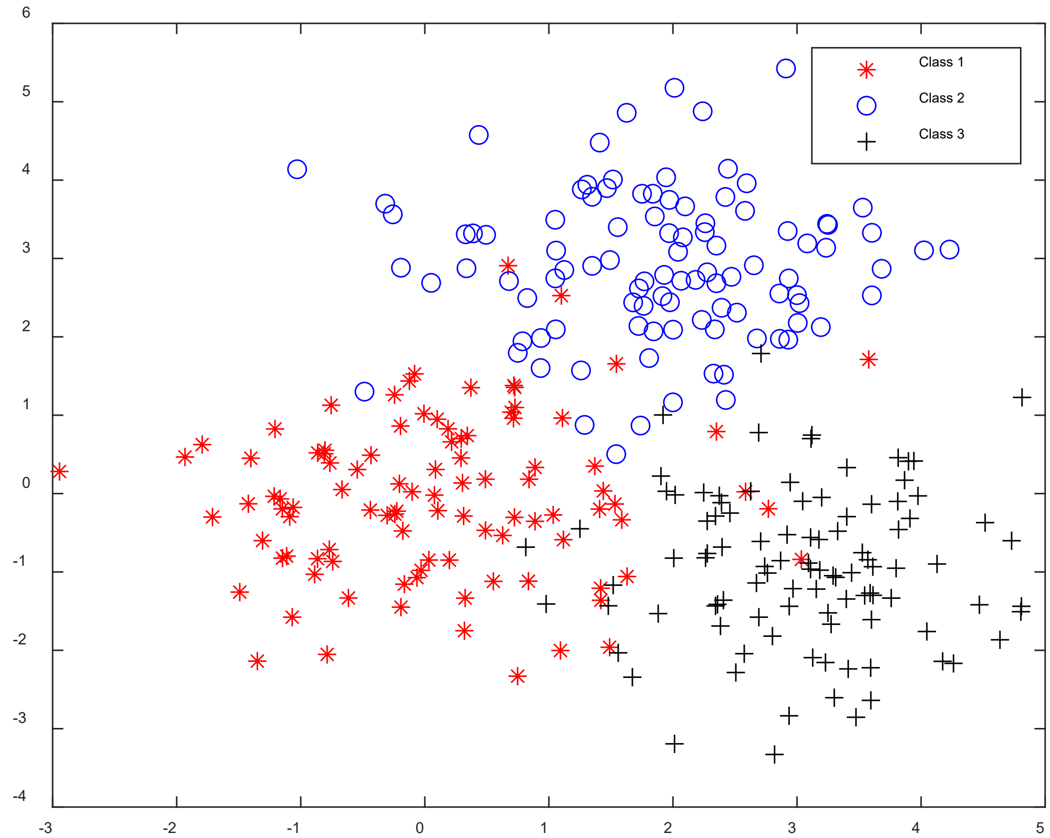
$$\mathbf{S}_W^{-1}\mathbf{S}_B\mathbf{w}_i = \lambda_i\mathbf{w}_i$$

This is a conventional eigenvalue problem, which again can be solved using some software toolbox such as Matlab.

Note, for a $c$ class problem, there are no more than $c-1$ eigenvectors.

## Example

Consider the following 3-class problem. dataset1 (attached) contain two files, where "data" contains the samples, and "label" contains the class label.



First, we calculate the sample mean vectors, and within-class and between-class scatter matrices:

$$\mathbf{m}_1 = \begin{bmatrix} 0.1083 \\ -0.0653 \end{bmatrix} \quad \mathbf{m}_2 = \begin{bmatrix} 1.8945 \\ 2.9026 \end{bmatrix} \quad \mathbf{m}_3 = \begin{bmatrix} 3.0515 \\ -0.935 \end{bmatrix}$$

$$\boldsymbol{S}_1 = \begin{bmatrix} 12640.28 & 1134.9 \\ 1134.9 & 10031.8 \end{bmatrix} \quad \boldsymbol{S}_2 = \begin{bmatrix} 10253.37 & -36.08 \\ -36.08 & 8880.35 \end{bmatrix}$$

$$\boldsymbol{S}_3 = \begin{bmatrix} 7250.18 & 47.82 \\ 47.82 & 9523.04 \end{bmatrix}$$

$$\boldsymbol{S}_W = (\boldsymbol{S}_1 + \boldsymbol{S}_2 + \boldsymbol{S}_3)/100 = \begin{bmatrix} 301.4383 & 11.4665 \\ 11.4665 & 284.352 \end{bmatrix}$$

**(Note: divided by 100 for scaling)**

$$\boldsymbol{S}_B = \begin{bmatrix} 439.7196 & -56.5992 \\ -56.5992 & 809.7615 \end{bmatrix}$$

Secondly, we find $\mathbf{w}$ by solving the following generalized eigenvalue problem:

$$\boldsymbol{S_B}\mathbf{w} = \lambda \boldsymbol{S}_w \mathbf{w}$$

which can be solved by using **M**atlab function “eig”:

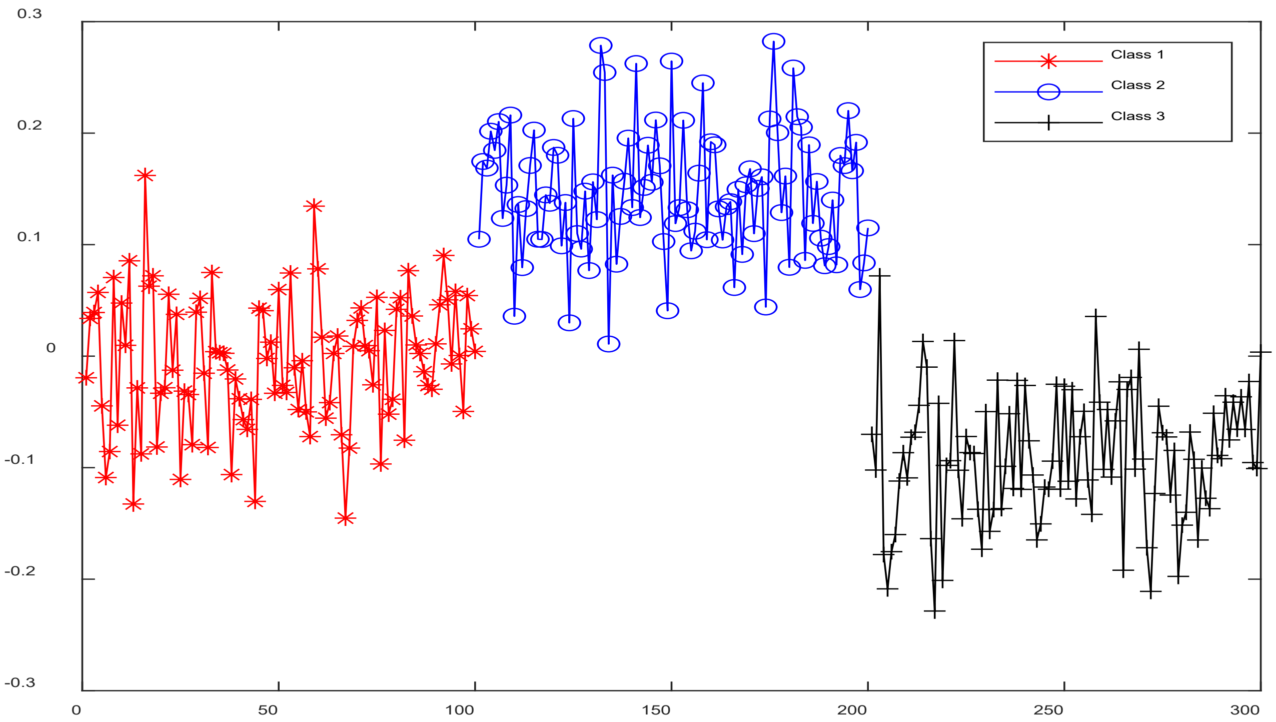
$$[\mathbf{W}, \boldsymbol{\Lambda}] = \mathrm{eig}(\boldsymbol{S}_B, \boldsymbol{S}_w)$$

Eigen values: $\boldsymbol{\Lambda} = \begin{bmatrix} 1.4155 & 0 \\ 0 & 2.9127 \end{bmatrix}$

Eigen vectors: $\mathbf{W} = \begin{bmatrix} -0.0564 & -0.012 \\ -0.0101 & 0.0585 \end{bmatrix}$

Let: $\mathbf{w}_1 = \begin{bmatrix} -0.012 \\ 0.0585 \end{bmatrix}$

Sample projection on $\mathbf{w}_1$:



For the projection of samples on $\mathbf{w}_1$, the mean values of the 3 classes are:

$$\widetilde{m}_{11} = -0.0051$$

$$\widetilde{m}_{12} = 0.1470$$

$$\widetilde{m}_{13} = -0.0913$$

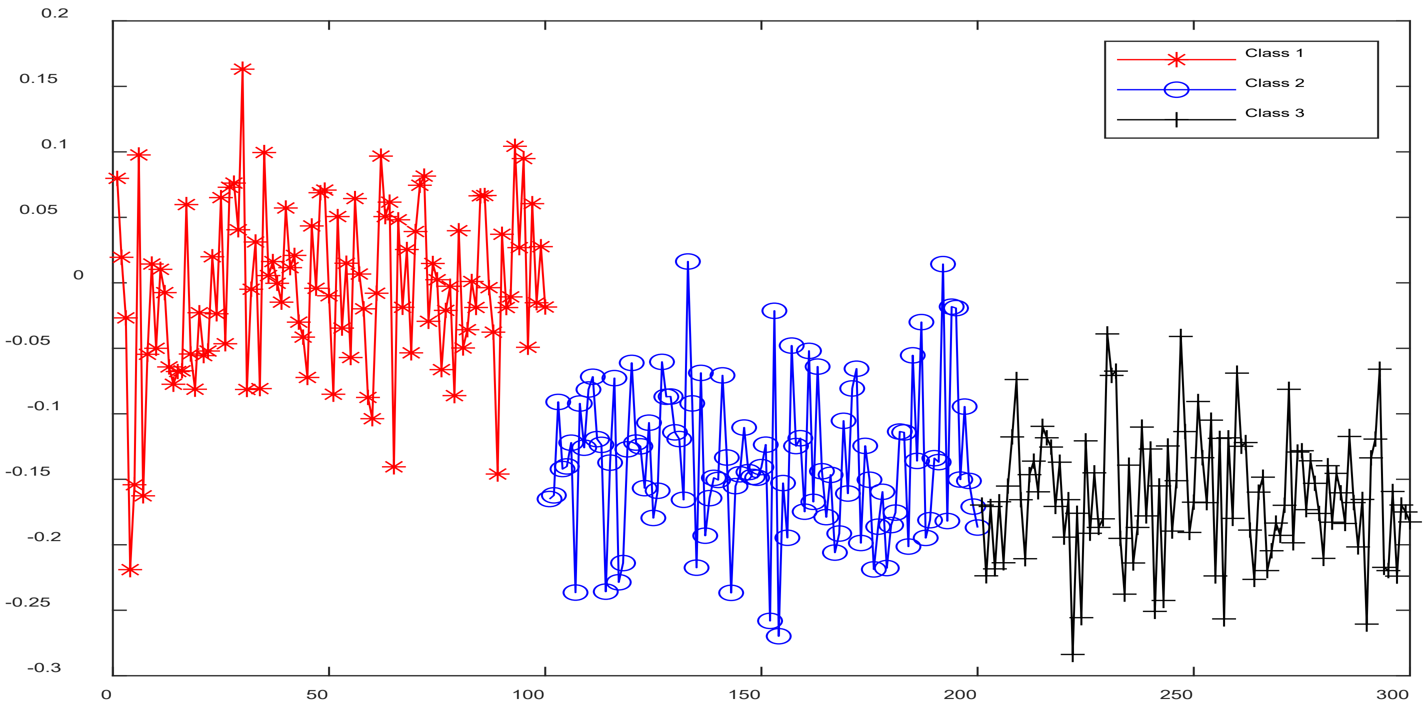
Samples in class 2 are well separated from samples in other two classes. In addition, samples in class 1 are closer to class 2. Therefore, we set:

$$w_{01} = -\frac{\widetilde{m}_{11} + \widetilde{m}_{12}}{2} = -0.0719$$

Let:

$$\mathbf{w}_2 = \begin{bmatrix} -0.0564 \\ -0.0101 \end{bmatrix}$$

Projection of samples on $w_2$:



For the projection of samples on $\mathbf{w}_2$, the mean values of the 3 classes are:

$$\widetilde{m}_{21} = -0.0054$$

$$\widetilde{m}_{22} = -0.1361$$

$$\widetilde{m}_{23} = -0.1626$$

Samples in class 1 are well separated from samples in other two classes. In addition, samples in class 2 are closer to class 1. Therefore, we set:

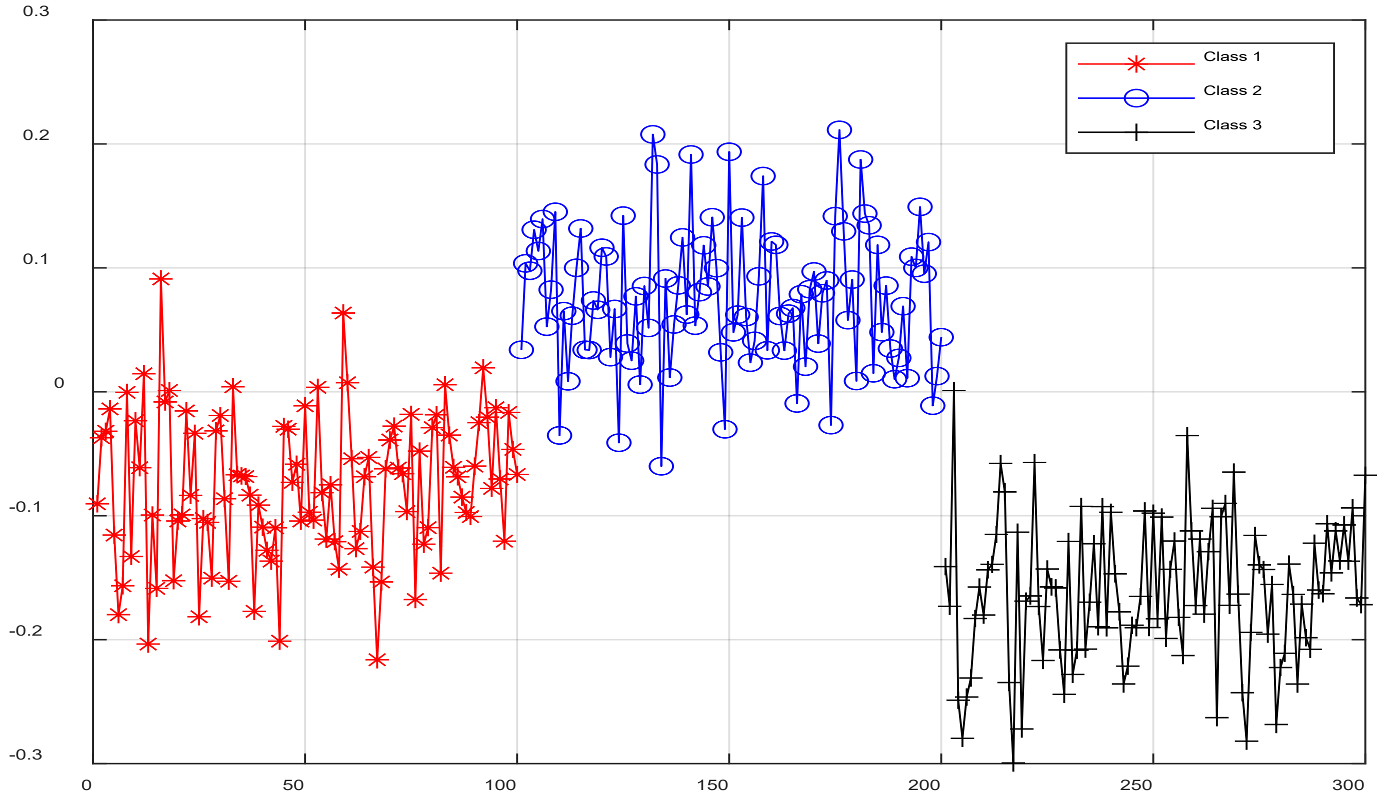
$$w_{02} = -\frac{\widetilde{m}_{21} + \widetilde{m}_{22}}{2} = 0.0708$$

We have the following discriminant functions:

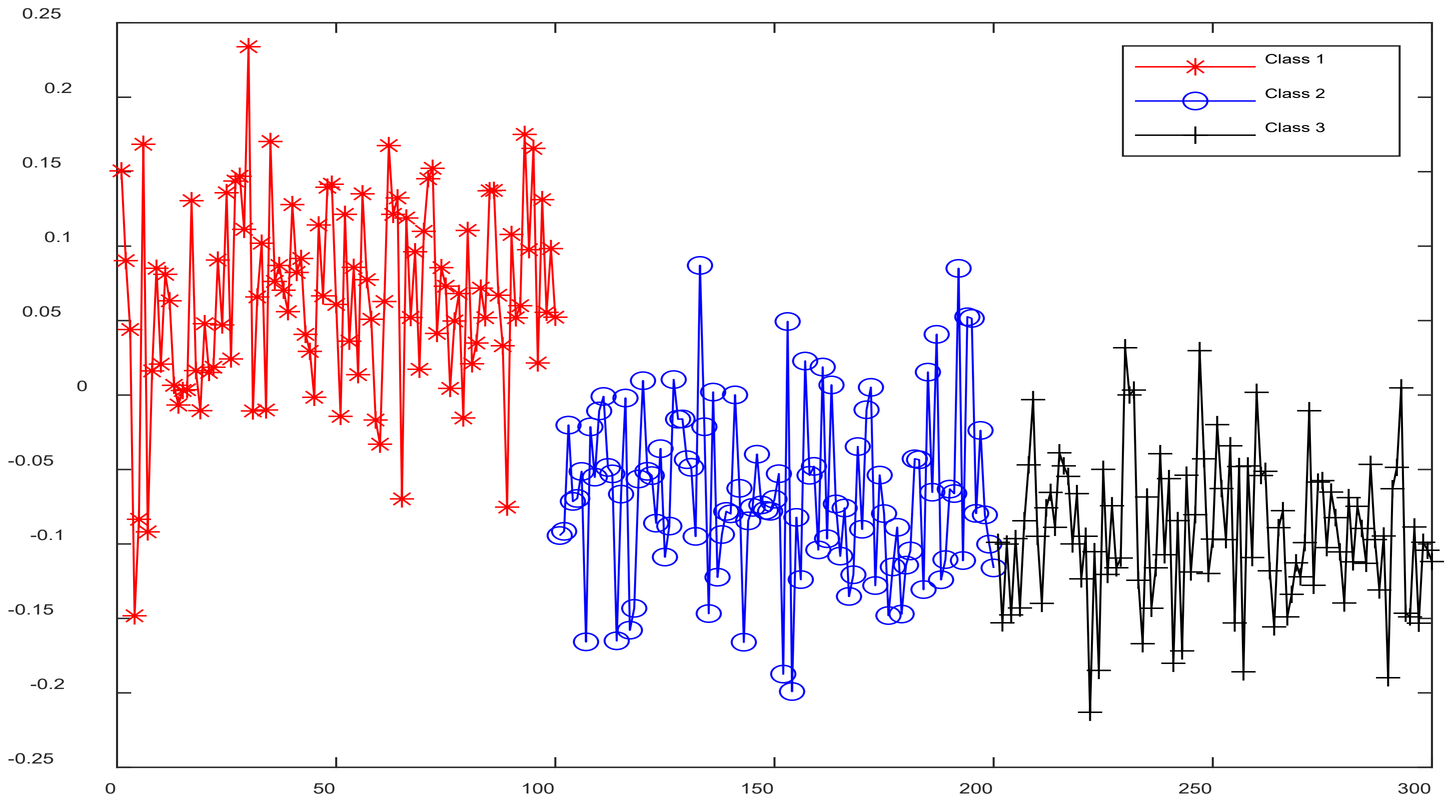
$$g_1(\mathbf{x}) = \mathbf{w}_1^T\mathbf{x} + w_{01} = \begin{bmatrix} -0.012 \\ 0.0585 \end{bmatrix}^T \mathbf{x} - 0.0719$$

$$g_2(\mathbf{x}) = \mathbf{w}_2^T\mathbf{x} + w_{02} = \begin{bmatrix} -0.0564 \\ -0.0101 \end{bmatrix}^T \mathbf{x} + 0.0708$$

# Discrimination function $g_1(x)$:



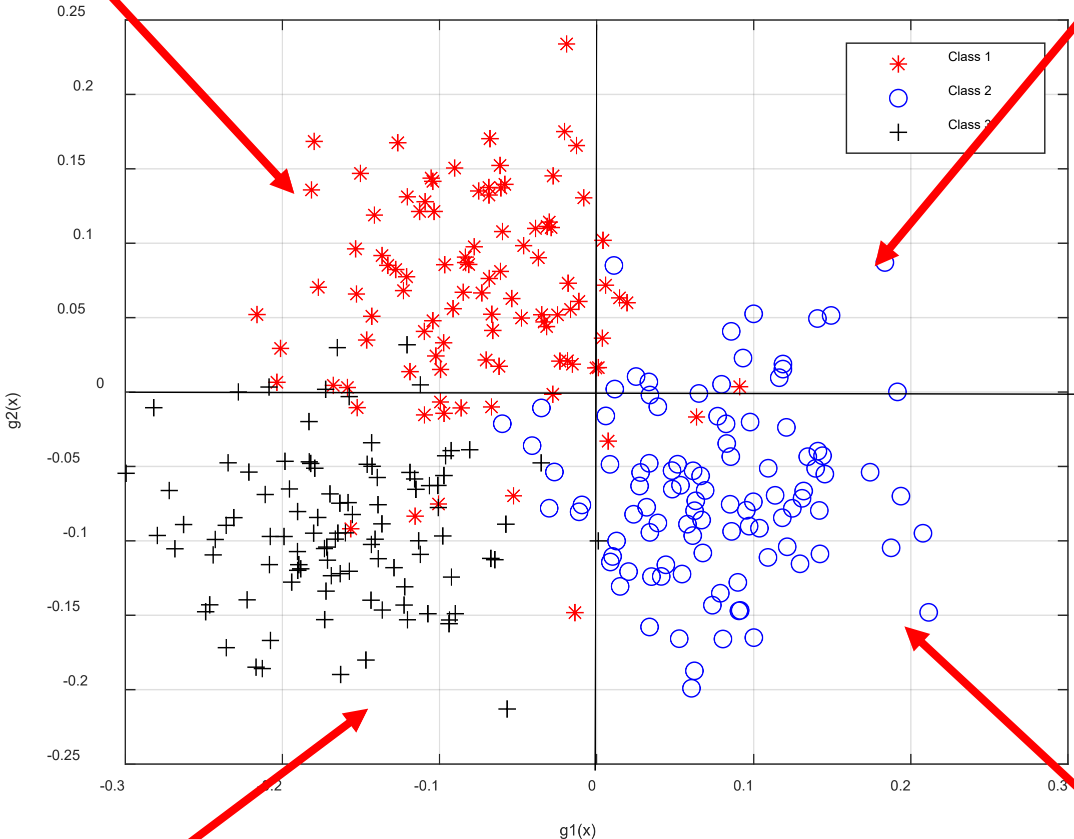
Discrimination function $g_2(x)$:



$g_2(\mathbf{x}) > 0, g_1(\mathbf{x}) < 0$

$g_1(\mathbf{x}) > 0, g_2(\mathbf{x}) > 0$

$g_1(\mathbf{x}) = 0$




$g_2(\mathbf{x}) = 0$

$g_1(\mathbf{x}) < 0, g_2(\mathbf{x}) < 0$

$g_1(\mathbf{x}) > 0, g_2(\mathbf{x}) < 0$

**Decision rules:**

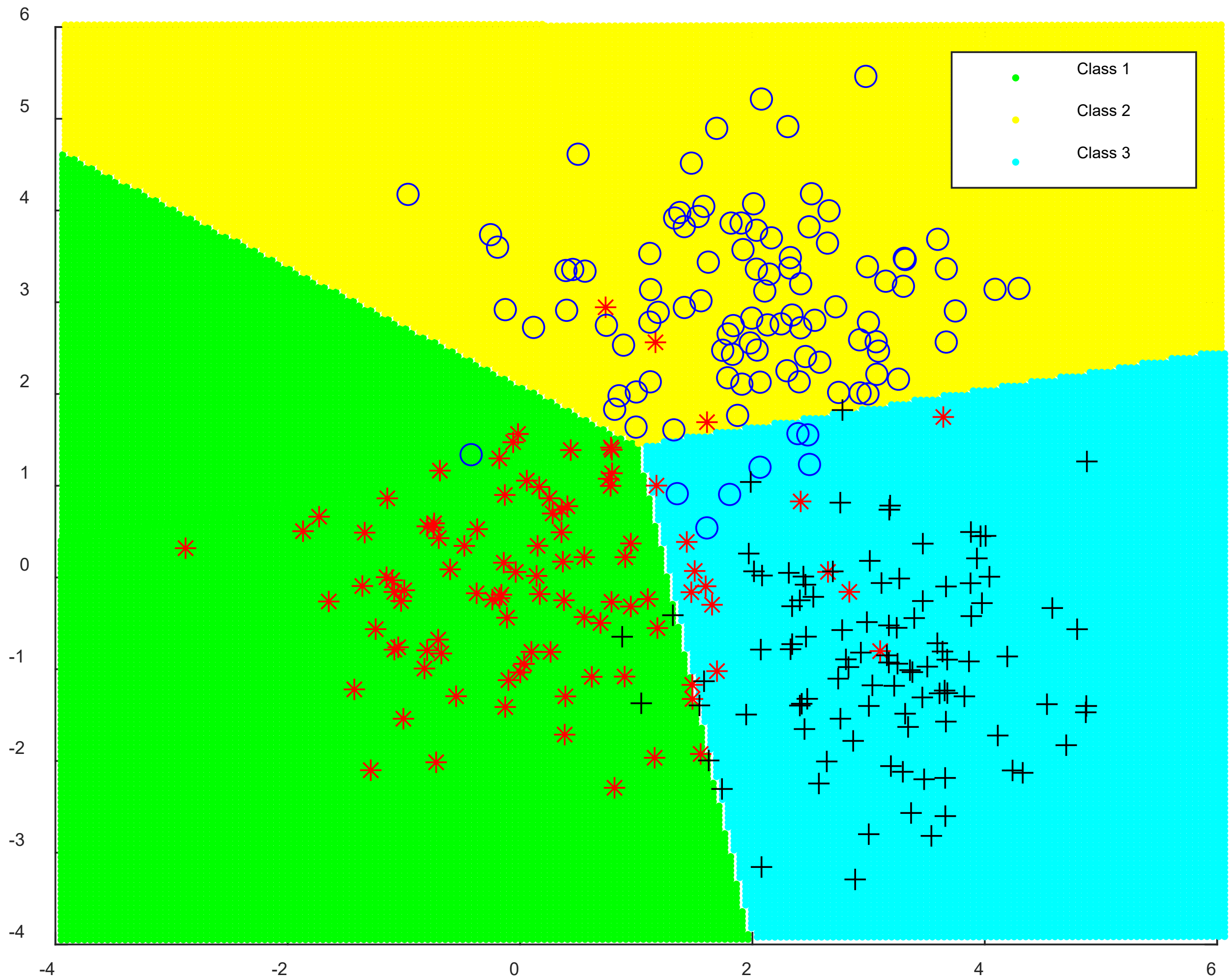
(1) If $g_1(\mathbf{x}) > 0$ and $g_2(\mathbf{x}) < 0$, decide $Class$ 2

(2) If $g_2(\mathbf{x}) > 0$ and $g_1(\mathbf{x}) < 0$, decide *Class 1*

(3) If $g_1(\mathbf{x}) < 0$ and $g_2(\mathbf{x}) < 0$, decide $Class$ 3

(4) If $g_1(\mathbf{x}) > 0$ and $g_2(\mathbf{x}) > 0$, decide $Class$ 2 if $g_1(\mathbf{x}) > g_2(\mathbf{x})$

decide $Class$ 1 if $g_2(\mathbf{x}) > g_1(\mathbf{x})$



Class 1
Class 2
Class 3
6
5
4
3
2
1
0
-1
-2
-3
-4
-4
-2
0
2
4
6

If rule (4) is not included, we have an **undecided region** as shown below: