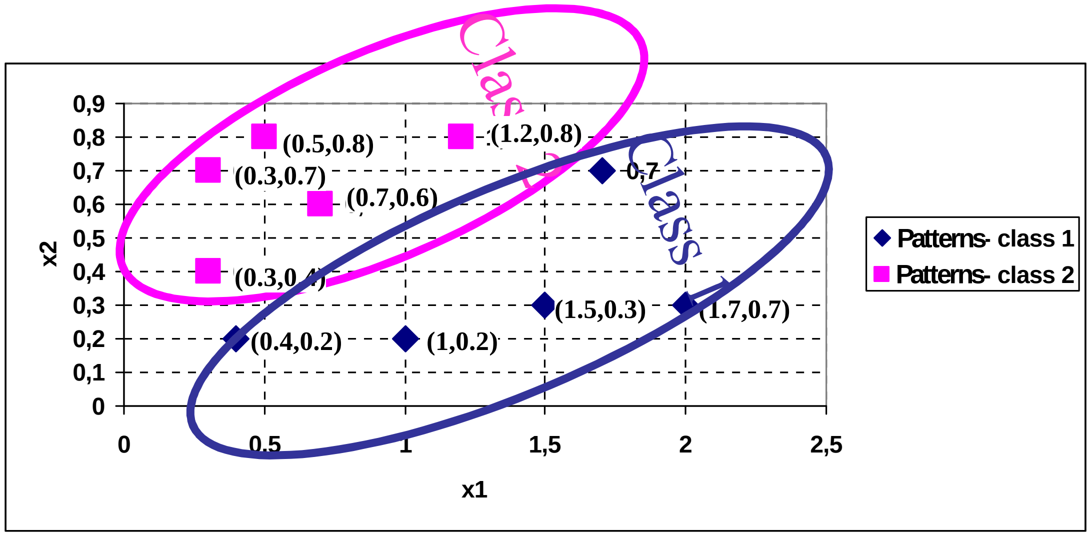
Discriminant analysis

The discriminant analysis aims at separation of patterns that belong to different classes. For example, if patterns are represented by points (each point is a pattern) then:

	Patterns – class 1	
	x11	x12
X1	0,4	0,2
X2	1	0,2
X3	2	0,3
X4	1,5	0,3
X5	1,7	0,7

	Patterns – class 2	
	x21	x22
X6	0,3	0,7
X7	0,5	0,8
X8	0,3	0,4
X9	0,7	0,6
X10	1,2	0,8

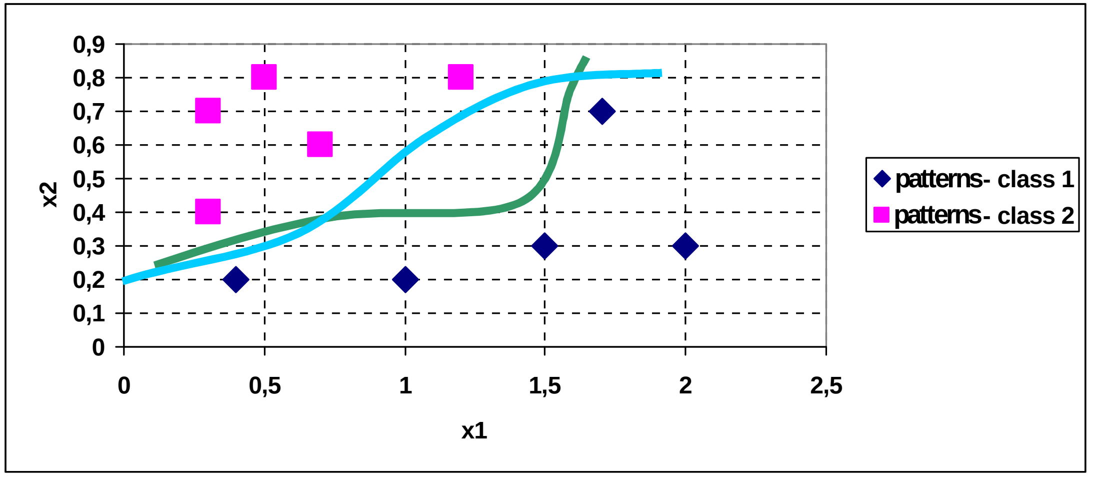


Classes are usually denoted by ω_1 and ω_2

The criterion of the separation is usually formulated as a function. Simultaneously, it can be simplified to two-class discrimination. Then the discrimination function is formulated as:

$$h(X) = \begin{cases} > k \Rightarrow X \in \Omega_1 \\ < k \Rightarrow X \in \Omega_2 \end{cases} \qquad k \text{ - constant}$$

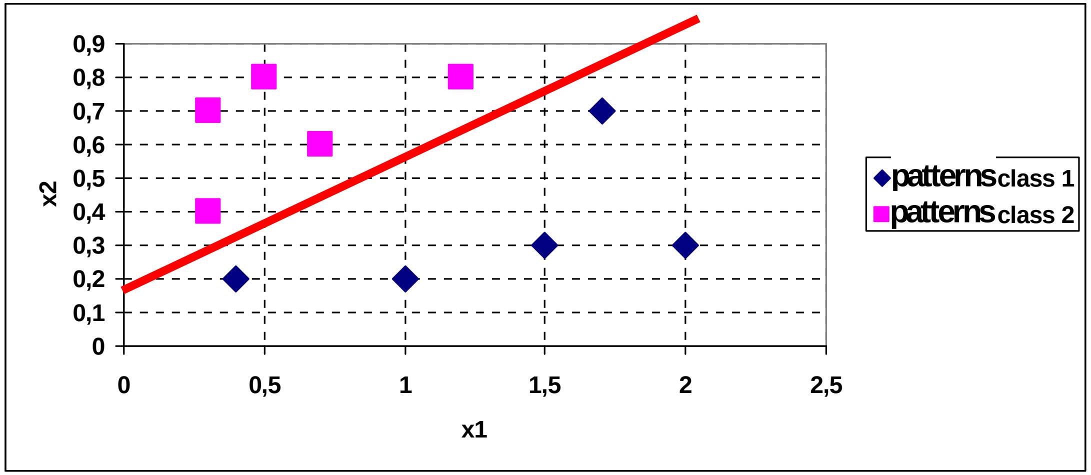
For $h(X) = k$ decision depend on us



If $h(X)$ separates two classes, then also $f(h(X))$ separates them if f is monotonous

$$g(X) = f(h(X)) = \begin{cases} > k' \Rightarrow X \in \Omega_1 \\ < k' \Rightarrow X \in \Omega_2 \end{cases}$$

Let us first solve the simplest task, i.e. lets separate two classes by a line.



$$x_2 = \omega_1 x_1 + \omega_0$$

$$\omega_1 x_1 - x_2 + \omega_0 = 0$$

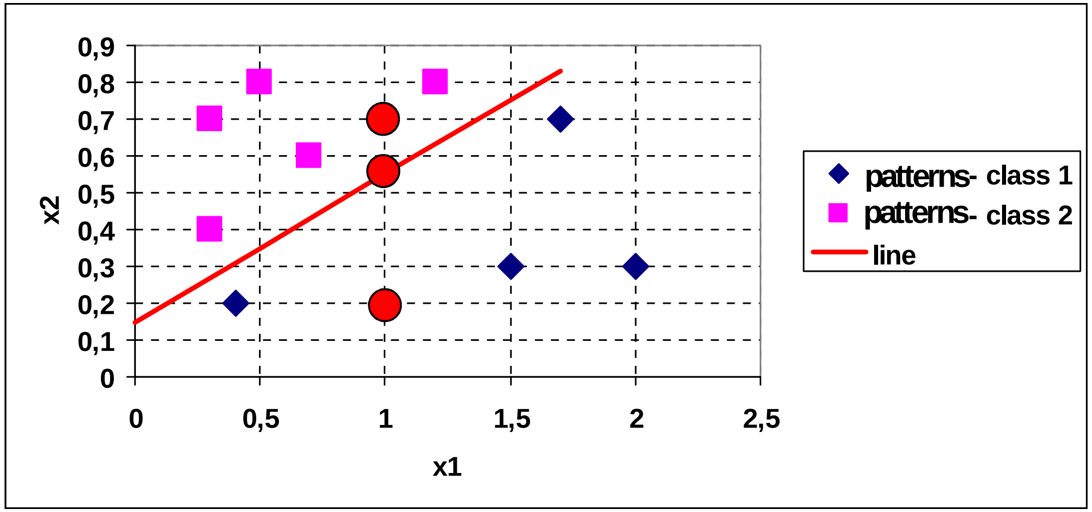
$$\omega_1 x_1 + \omega_2 x_2 + \omega_0 = 0$$

The linear discriminant function is defined as:

$$[\omega_1 \quad \omega_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \omega_0 = \omega_1 x_1 + \omega_2 x_2 + \omega_0 =$$

$$g(X) = \varpi^T \tilde{X} + \omega_0 = \boldsymbol{\omega}^T X$$

$$\omega_1 x_1 + \omega_2 x_2 + \omega_0 \cdot 1 = \overset{\boldsymbol{\omega}^T}{[\omega_1 \quad \omega_2 \quad \omega_0]} \overset{X}{\begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}}$$



$$x_2 = 0.4x_1 + 0.15$$

$$0.4x_1 - x_2 + 0.15 = 0$$

$$\omega_1 = 0.4 \quad \omega_2 = -1 \quad \omega_0 = 0.15$$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

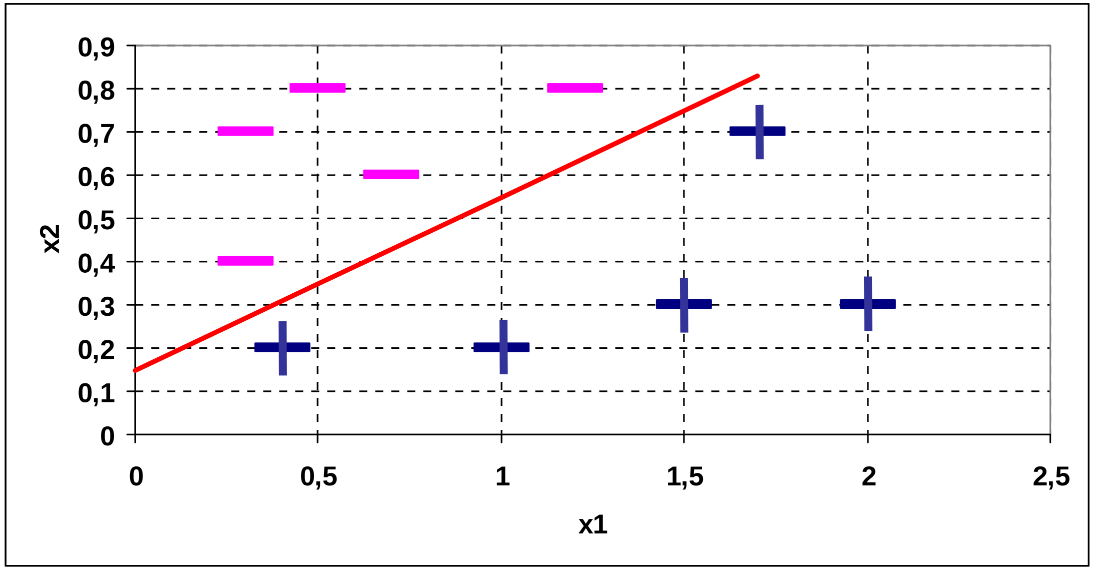
$$\begin{bmatrix} \omega_1 & \omega_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \omega_0 = \omega_1 x_1 + \omega_2 x_2 + \omega_0 =$$

$$\omega_1 x_1 + \omega_2 x_2 + \omega_0 \cdot 1 = \begin{bmatrix} \omega_1 & \omega_2 & \omega_0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

			X
			1
			0,55
ω^T			1
0,4	-1	0,15	0

			X
			1
			0,2
ω^T			1
0,4	-1	0,15	0,35

			X
			1
			0,7
ω^T			1
0,4	-1	0,15	-0,15



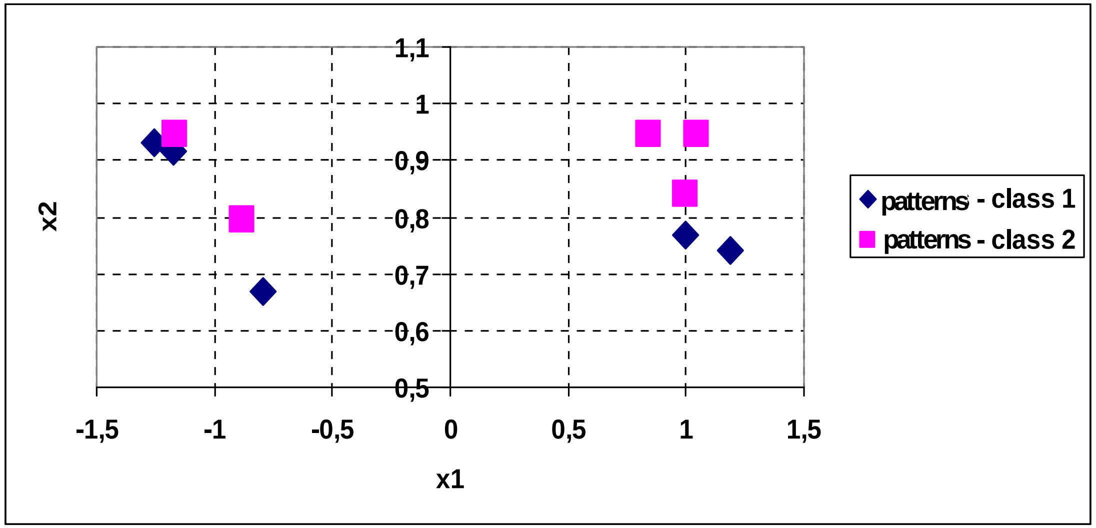
Now, we have a simple criterion:

$$\boldsymbol{\omega}^T X \begin{cases} > 0 \Rightarrow X \in \Omega_1 \\ < 0 \Rightarrow X \in \Omega_2 \end{cases}$$

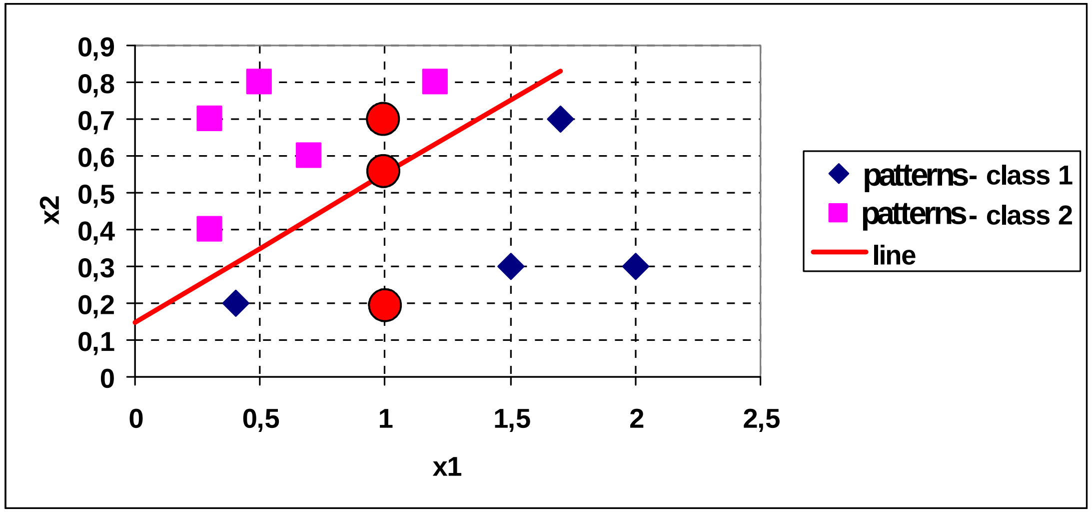
In case of equality we decide, as previously

Sometimes data that at the first glance are not lineary separable can be separeted after transformation.

$$g(X) = f(h(X)) = \begin{cases} > k' \Rightarrow X \in \Omega_1 \\ < k' \Rightarrow X \in \Omega_2 \end{cases}$$



$f(x)=x^2$



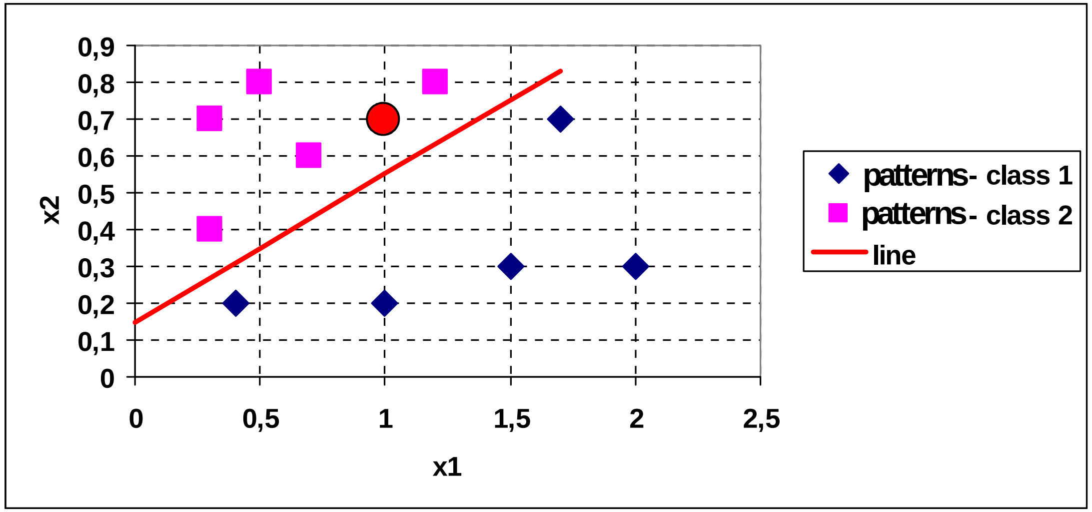
			X
			1
			0,55
ω^T			1
0,4	-1	0,15	0

			X
			1
			0,2
ω^T			1
0,4	-1	0,15	0,35

We can multiply patterns of one class by -1

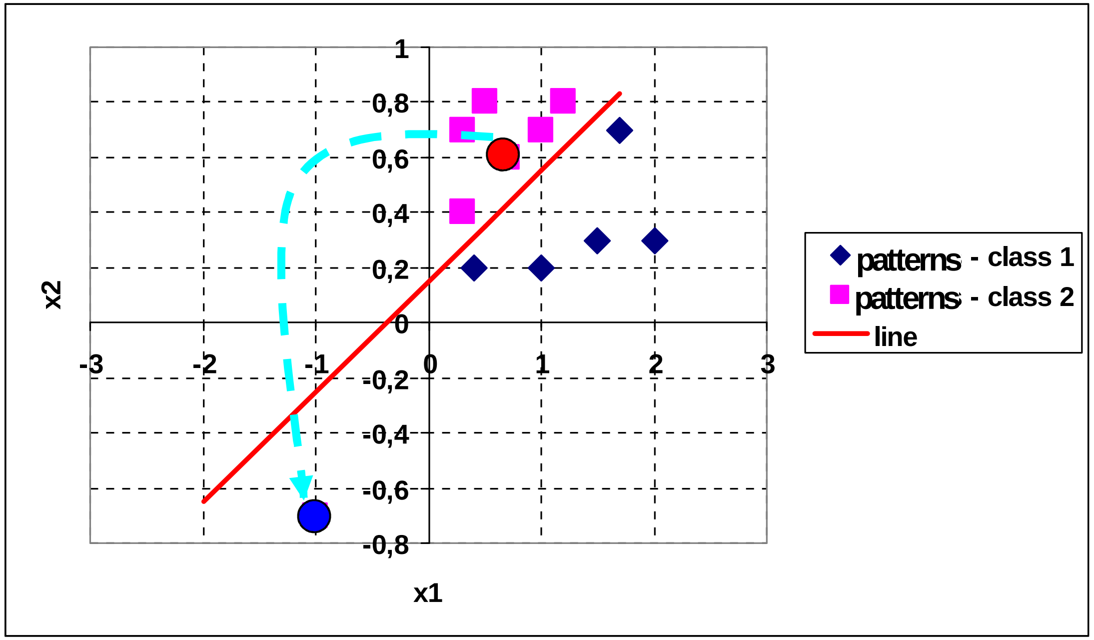
			X
			-1
			-0,7
			-1
0,4	-1	0,15	0,15

			X
			1
			0,7
ω^T			1
0,4	-1	0,15	-0,15



			X
			1
ω^T			0,7
			1
0,4	-1	0,15	-0,15

			X
			-1
			-0,7
			-1
0,4	-1	0,15	0,15



Thus, we can make the criterion:

$$\boldsymbol{\omega}^T X \begin{cases} > 0 \Rightarrow X \in \Omega_1 \\ < 0 \Rightarrow X \in \Omega_2 \end{cases}$$

even simpler. If in the set $\{X\}$ we will multiply all samples that belong to Ω_2 by –1, we have:

$$\mathbf{y}_i^T = \begin{cases} X_i^T \text{ if } X_i \in \Omega_1 \\ -X_i^T \text{ if } X_i \in \Omega_2 \end{cases}$$

and we can formulate criterion to find the line that separates the two classes:

$$\boldsymbol{\omega}^T \mathbf{y} > 0$$

The perceptron criterion

The percepcion criterion is based on misclassified patterns. Let us define Y – the set of misclassified patterns.

$$Y = \left\{ y_i , \omega^T y_i < 0 \right\}$$

Then the classification criterion is:

$$J_p = \sum_{y_i \in Y} \left(-\omega^T y_i \right)$$

The ω should be changed in such a way that $J_p \to 0$. The change of J_p related to change of ω is:

$$\frac{\partial J_p}{\partial \omega} = \sum_{y_i \in Y} \left(-y_i \right)$$

i.e. the sum of misclassified patterns.

Thus, ω should be updated by means of the sum of misclassified patterns

$$\omega_{k+1} = \omega_k + \rho_k \sum_{\mathrm{y}_i \in Y} \mathrm{y}_i$$

where ρ_k determines how fast ω is modified.

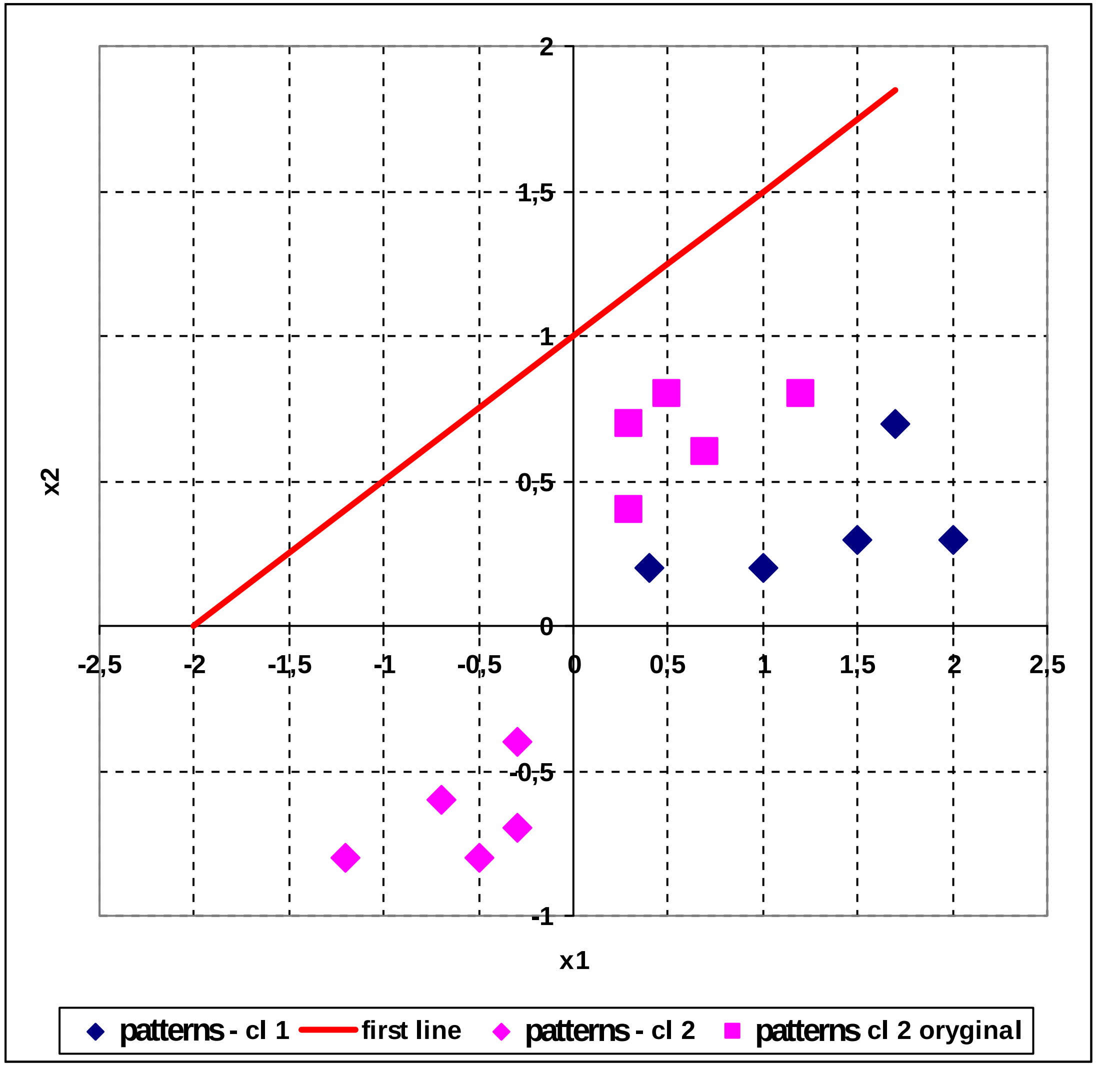
patterns - class 1 and 2			patterns - class 1 and 2 afer transformation *(-1)		
x1	x2		x1	x2	
0,4	0,2	Y1	0,4	0,2	1
1	0,2	Y2	1	0,2	1
2	0,3	Y3	2	0,3	1
1,5	0,3	Y4	1,5	0,3	1
1,7	0,7	Y5	1,7	0,7	1
0,3	0,7	Y6	-0,3	-0,7	-1
0,5	0,8	Y7	-0,5	-0,8	-1
0,3	0,4	Y8	-0,3	-0,4	-1
0,7	0,6	Y9	-0,7	-0,6	-1
1,2	0,8	Y10	-1,2	-0,8	-1

*

first line				
		om1	om2	om0
		1	-2	2
	x2=	0,5	*x1+	1

$\omega^T Y$

primary classification	
2	
2,6	
3,4	
2,9	
2,3	
-0,9	miscl
-0,9	miscl
-1,5	miscl
-1,5	miscl
-1,6	miscl



l 2	**patterns - class 1 and 2 afer transformation *(-1)**			primary classification	
	x1	x2			
Y1	0,4	0,2	1	2	
Y2	1	0,2	1	2,6	
Y3	2	0,3	1	3,4	
Y4	1,5	0,3	1	2,9	
Y5	1,7	0,7	1	2,3	
Y6	-0,3	-0,7	-1	-0,9	miscl
Y7	-0,5	-0,8	-1	-0,9	miscl
Y8	-0,3	-0,4	-1	-1,5	miscl
Y9	-0,7	-0,6	-1	-1,5	miscl
Y10	-1,2	-0,8	-1	-1,6	miscl

	x1	x2	
Sum of misclassified patterns:	-3	-3,3	-5

Previous ω:

om1	om2	om0
1	-2	2

$$\omega_{k+1} = \omega_k + \rho_k \sum_{y_i \in Y} y_i$$

ro=	**0,25**

			om1	om2	om0
			0,25	-2,83	0,75
second line	x2=		0,088	*x1+	0,265

l 2	patterns - class 1 and 2 afer transformation *(-1)			primary classification	
	x1	x2			
Y1	0,4	0,2	1	2	
Y2	1	0,2	1	2,6	
Y3	2	0,3	1	3,4	
Y4	1,5	0,3	1	2,9	
Y5	1,7	0,7	1	2,3	
Y6	-0,3	-0,7	-1	-0,9	miscl
Y7	-0,5	-0,8	-1	-0,9	miscl
Y8	-0,3	-0,4	-1	-1,5	miscl
Y9	-0,7	-0,6	-1	-1,5	miscl
Y10	-1,2	-0,8	-1	-1,6	miscl

ro=	0,25

om1	om2	om0
1	-2	2

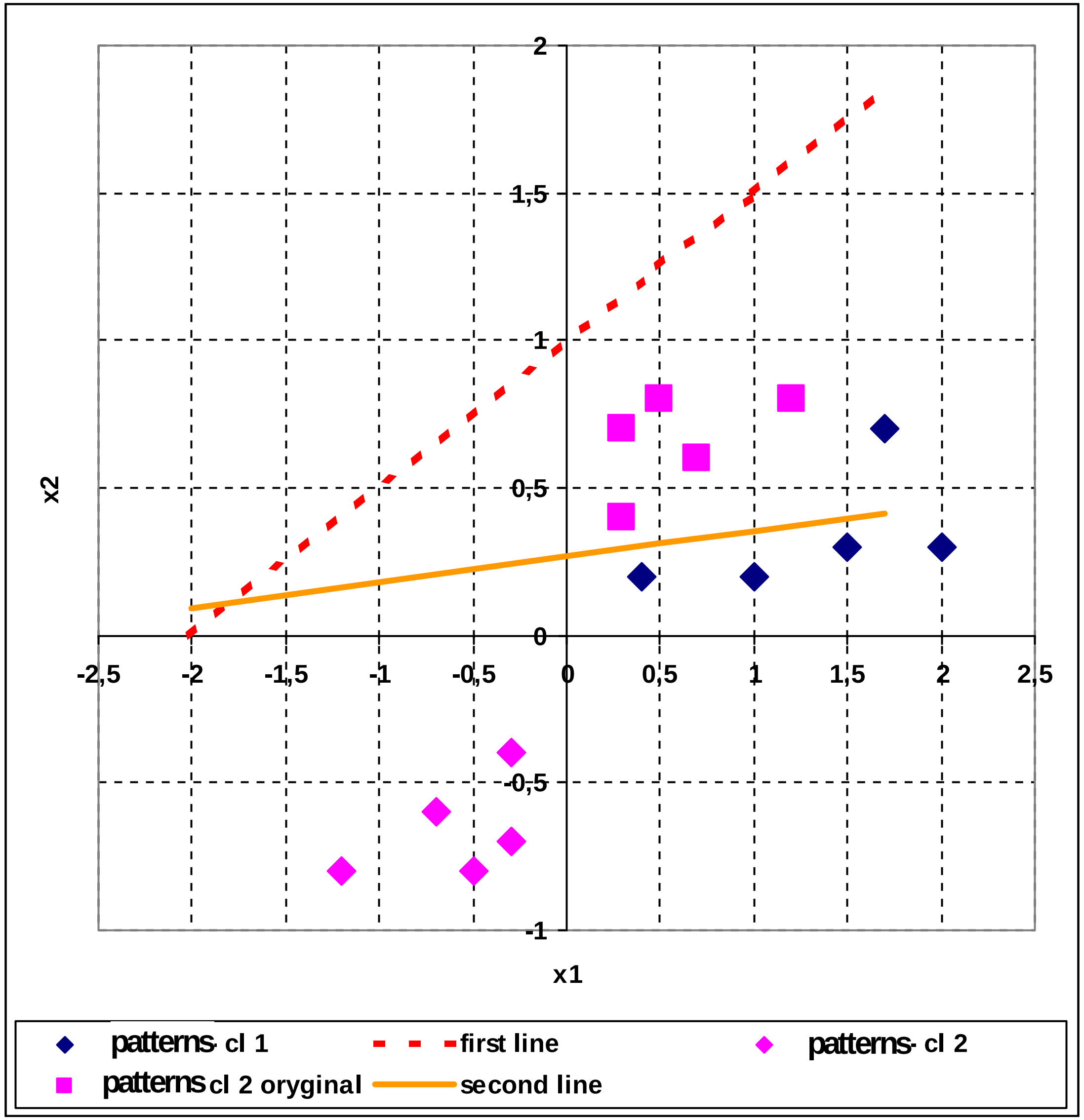
Previous:

$$\omega_{k+1} = \omega_k + \rho_k \sum_{y_i \in Y} y_i$$

			om1	om2	om0
			0,25	-2,825	0,75
second line		x2=	0,088496	*x1+	0,2654867

$$\omega^T Y$$

0,285	
0,435	
0,4025	
0,2775	
-0,8025	miscl
1,1525	
1,385	
0,305	
0,77	
1,21	



l 2	**patterns - class 1 and 2 afer transformation *(-1)**			secondary classification	
	x1	x2			
Y1	0,4	0,2	1	0,285	
Y2	1	0,2	1	0,435	
Y3	2	0,3	1	0,4025	
Y4	1,5	0,3	1	0,2775	
Y5	1,7	0,7	1	-0,8025	miscl
Y6	-0,3	-0,7	-1	1,1525	
Y7	-0,5	-0,8	-1	1,385	
Y8	-0,3	-0,4	-1	0,305	
Y9	-0,7	-0,6	-1	0,77	
Y10	-1,2	-0,8	-1	1,21	

		x1	**x2**	
sum of misclassified patterns:		1,7	0,7	1,00000

ro=	**0,3**

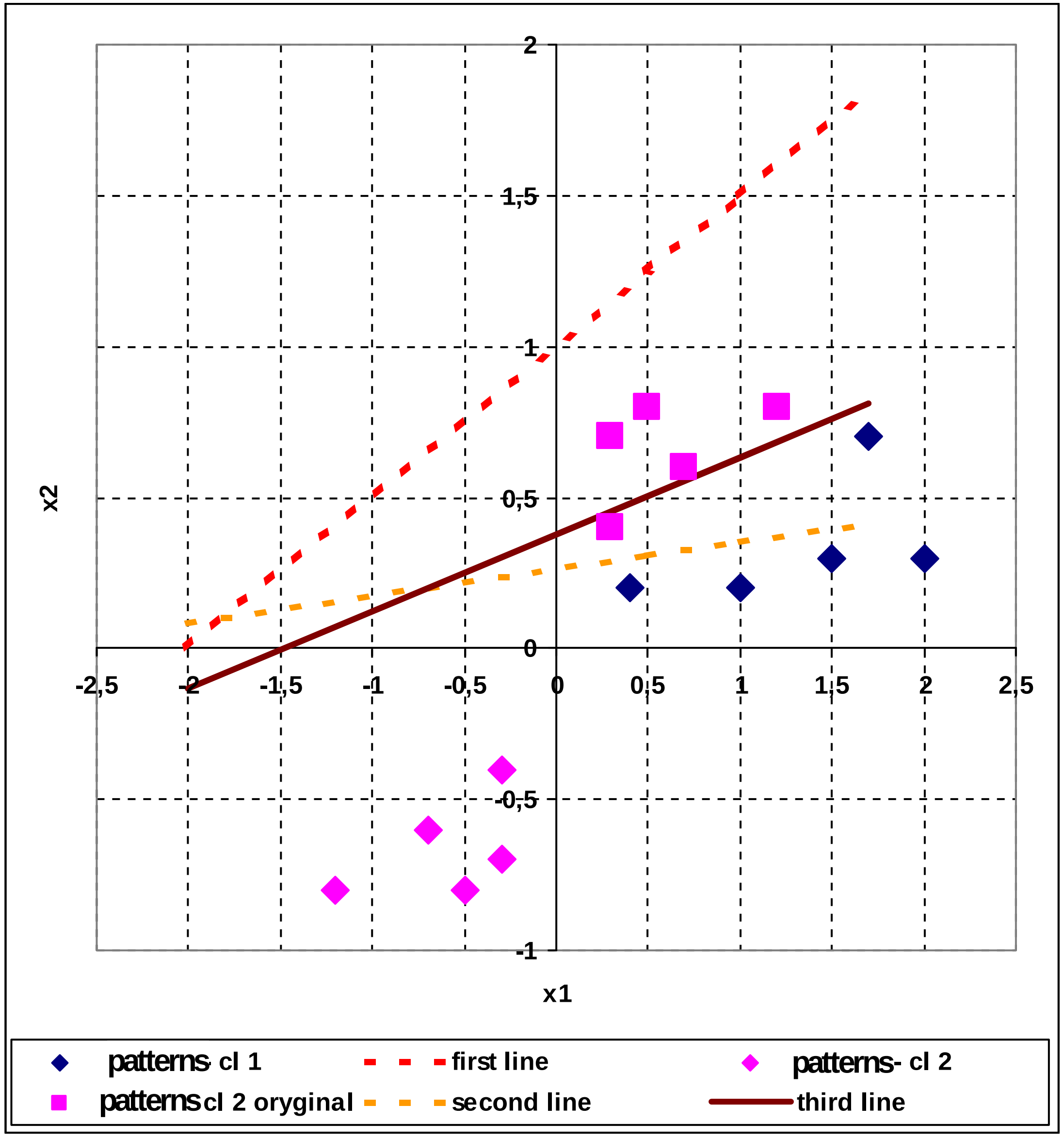
Previous:

om1	om2	om0
0,25	-2,825	0,75

			om1	om2	om0
			0,675	-2,65	1
third line		x2=	0,254717	*x1+	0,3773585

$$\omega^T Y$$

0,74	
1,145	
1,555	
1,2175	
0,2925	
0,6525	
0,7825	
-0,1425	miscl
0,1175	
0,31	



l 2	patterns - class 1 and 2 afer transformation *(-1)				
	x1	x2			
Y1	0,4	0,2	1	0,74	
Y2	1	0,2	1	1,145	
Y3	2	0,3	1	1,555	
Y4	1,5	0,3	1	1,2175	
Y5	1,7	0,7	1	0,2925	
Y6	-0,3	-0,7	-1	0,6525	
Y7	-0,5	-0,8	-1	0,7825	
Y8	-0,3	-0,4	-1	-0,1425	miscl
Y9	-0,7	-0,6	-1	0,1175	
Y10	-1,2	-0,8	-1	0,31	

			x1	x2	
sum of misclassified patterns:			-0,3	-0,4	-1

ro=	0,25

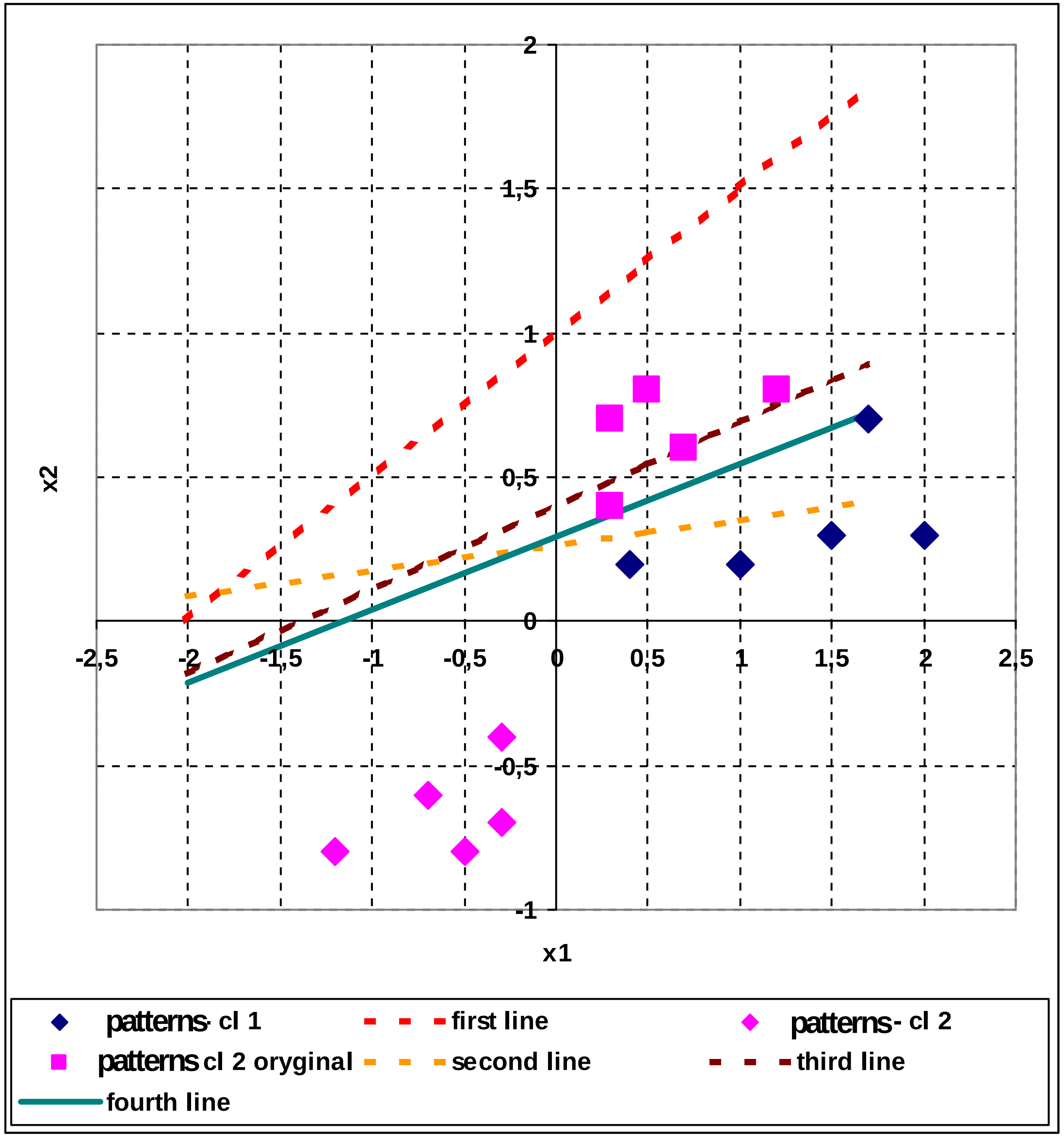
Previous:

om1	om2	om0
0,76	-2,615	1,05

			om1	om2	om0
			0,685	-2,715	0,8
fourth line		x2=	0,252302	*x1+	0,2946593

$$\omega^T Y$$

0,531
0,942
1,3555
1,013
0,064
0,895
1,0295
0,0805
0,3495
0,55



Variants

1) However, we may decide to change ω_k by means of a single misclassified pattern. Then:

$$\omega_{k+1} = \omega_k + \rho_k y_i$$

2) The ρ_k parameter could be also chosen in such a way that:

$$\omega_{k+1}^T y_i > 0$$

This means that the pattern y_i which was misclassified in the *i*-th iteration (because $\omega_k y_i < 0$) will be correctly classified in the *i*+1-th iteration ($\omega_{k+1} y_i > 0$). It will occur if:

$$\rho_k > \frac{\left|\omega_{k+1}^T y_i\right|}{|y_i|^2}$$

$$\omega_{k+1} = \omega_k + \rho_k y_i$$

$$\omega_{k+1}{}^T y_i = (\omega_k + \rho_k y_i)^T y_i = \omega_k{}^T y_i + \rho_k y_i{}^T y_i$$

$$\omega_{k+1}{}^T y_i = \omega_k{}^T y_i + \rho_k |y_i|^2$$

$$\omega_{k+1}{}^T y_i - \rho_k |y_i|^2 = \omega_k{}^T y_i < 0$$

$$\omega_{k+1}{}^T y_i - \rho_k |y_i|^2 < 0$$

$$\omega_{k+1}{}^T y_i < \rho_k |y_i|^2$$

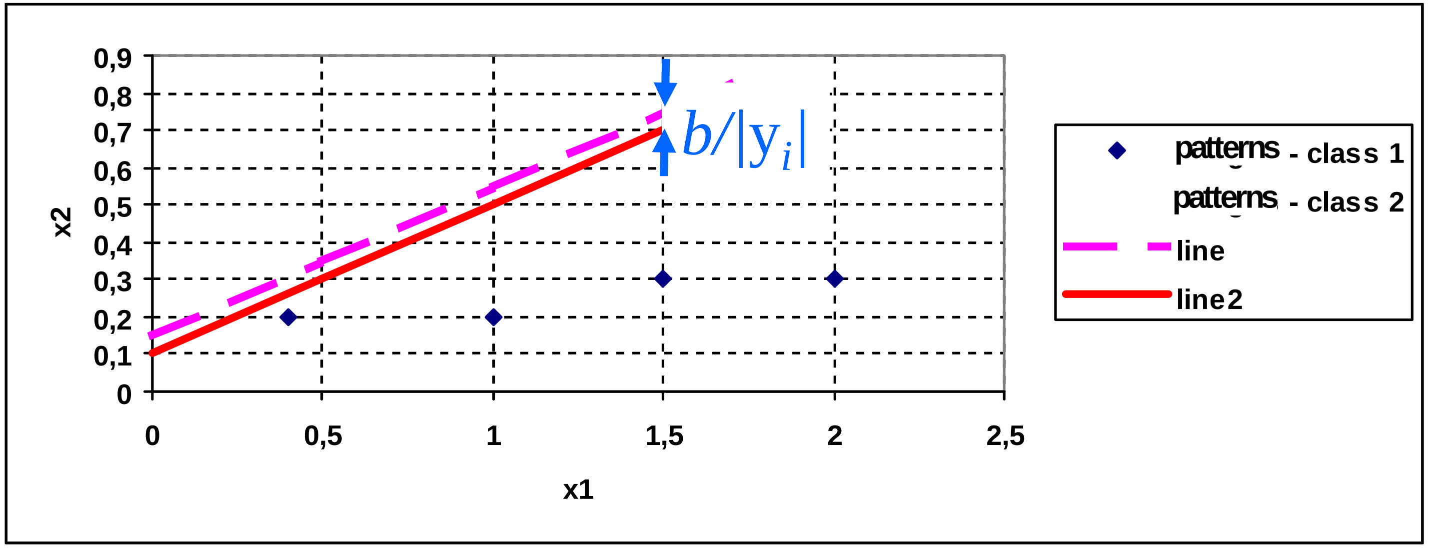
$$\rho_k > \frac{\omega_{k+1}{}^T y_i}{|y_i|^2}$$

Variants

3) The ω parameter is modified whenever „situation becomes dangerous”, i.e.:

$$\omega^T y_i \leq b, b > 0$$

This means that „a safety margin” is added to the discrimination line – it is shifted.



Relaxation algorithm

Relaxation algorithm minimizes the criterion:

$$J_R = \frac{1}{2} \sum_{y_i \in Y} \frac{\left(\omega^T y_i - b\right)^2}{\left|y_i\right|^2}$$

$$Y = \left\{ y_i, \omega^T y_i \le b \right\}$$

Weights are updated according to all (1) or only one (2) misclassified pattern(s):

$$1) \quad \omega_{k+1} = \omega_k + \rho_k \sum_{y_i \in Y} \frac{b - \omega^T y_i}{\left|y_i\right|^2} y_i$$

$$2) \quad \omega_{k+1} = \omega_k + \rho_k \frac{b - \omega^T y_i}{\left|y_i\right|^2} y_i$$

l 2	patterns - class 1 and 2 afer transformation *(-1)		
	x1	x2	
Y1	0,4	0,2	1
Y2	1	0,2	1
Y3	2	0,3	1
Y4	1,5	0,3	1
Y5	1,7	0,7	1
Y6	-0,3	-0,7	-1
Y7	-0,5	-0,8	-1
Y8	-0,3	-0,4	-1
Y9	-0,7	-0,6	-1
Y10	-1,2	-0,8	-1

b= 0,5

$\omega^T Y$

2	
2,6	
3,4	
2,9	
2,3	
-0,9	miscl
-0,9	miscl
-1,5	miscl
-1,5	miscl
-1,6	miscl

first line					
			om1	om2	om0
			1	-2	2
		x2=	0,5	*x1+	1

ro=	0,25

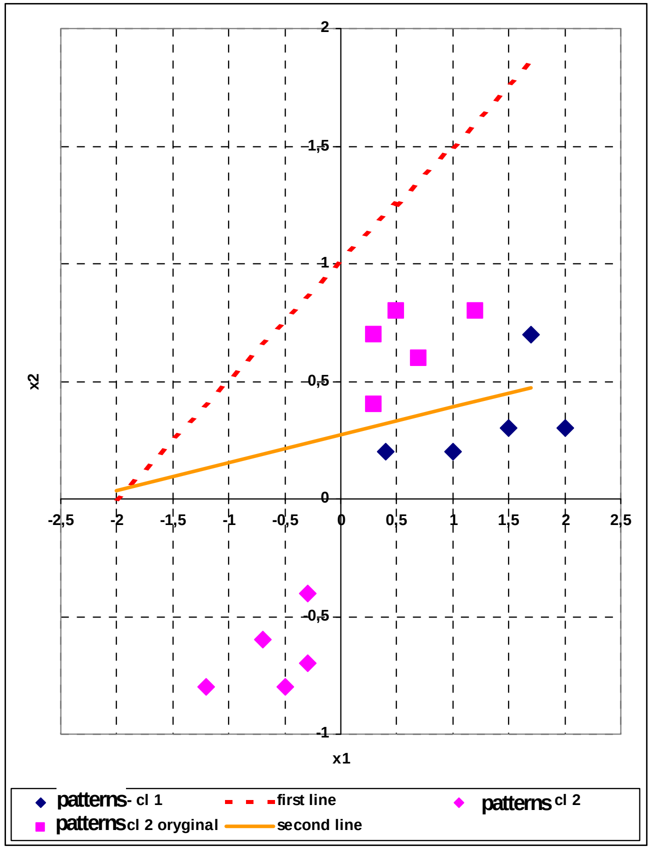
COEF

+

$$\omega_{k+1} = \omega_k + \rho_k \sum_{y_i \in Y} \frac{b - \omega^T y_i}{|y_i|^2} y_i$$

COEF	PATTERN			COEF*PAT*RO		
0,8860759	-0,3	-0,7	-1	-0,06646	-0,15506	-0,22152
0,7407407	-0,5	-0,8	-1	-0,09259	-0,14815	-0,18519
1,6	-0,3	-0,4	-1	-0,12	-0,16	-0,4
1,0810811	-0,7	-0,6	-1	-0,18919	-0,16216	-0,27027
0,6818182	-1,2	-0,8	-1	-0,20455	-0,13636	-0,17045
			SUM	-0,67278	-0,76174	-1,24743

			om1	om2	om0
			0,327217	-2,76174	0,752571
second line		x2=	0,118482	*x1+	0,2724991



Fisher’s approach to linear dicriminant analysis

Fisher has taken into consideration similarity of patterns inside a class vs. dissimilarity between classes. Thus the criterion:

$$J_F = \frac{\left|\omega^T(\mathrm{m}_1 - \mathrm{m}_2)\right|^2}{\omega^T \mathrm{S}_w \omega}$$

should be maximum. Where:

mean pattern of the *j*-th class

$$\mathrm{m}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} \mathrm{x}_{ij}$$

x_{ij} – *i*-th pattern of the *j*-th class

$$\mathrm{S}_w = \frac{1}{n_1 + n_2 - 2}\left(n_1\mathrm{S}_1 + n_2\mathrm{S}_2\right)$$

$$\mathrm{S}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\left(\mathrm{x}_{ij} - \mathrm{m}_j\right)\left(\mathrm{x}_{ij} - \mathrm{m}_j\right)^T$$

It can be prooved that we can take:

$$\omega = S_w^{-1}(m_1 - m_2)$$

The x pattern is assigned to class *k* if:

$$\left|\omega^T x - \omega^T m_k\right| < \left|\omega^T x - \omega^T m_j\right|, \quad k \neq j$$

Coefficients of the classification line

$$\left|\omega^T x - \omega^T m_k\right| = \left|\omega^T x - \omega^T m_j\right|$$

1) $\cancel{\omega^T x} - \omega^T m_k = \cancel{\omega^T x} - \omega^T m_j$

$$\omega^T m_k = \omega^T m_j$$

2) $\omega^T x - \omega^T m_k = -\omega^T x + \omega^T m_j$

$$2\omega^T x = \omega^T \left(m_k + m_j\right)$$

$$\omega_1 x_1 + \omega_2 x_2 = \frac{1}{2}\omega^T \left(m_k + m_j\right)$$

$$x_2 = -\frac{\omega_1}{\omega_2} x_1 + \frac{1}{2\omega_2}\omega^T \left(m_k + m_j\right)$$

		x1	x2			
class1	X1T	0,4	0,2	m1T	1,32	0,34
	X2T	1	0,2			
	X3T	2	0,3			
	X4T	1,5	0,3			
	X5T	1,7	0,7			
class2	X6T	0,3	0,7	m2T	0,6	0,66
	X7T	0,5	0,8			
	X8T	0,3	0,4			
	X9T	0,7	0,6			
	X10T	1,2	0,8			

$$\mathrm{S}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \left(\mathrm{x}_{ij} - \mathrm{m}_j\right)\left(\mathrm{x}_{ij} - \mathrm{m}_j\right)^T$$

$$\left(\mathrm{x}_{ij} - \mathrm{m}_j\right)^T$$

-0,92	-0,14
-0,32	-0,14
0,68	-0,04
0,18	-0,04
0,38	0,36
-0,3	0,04
-0,1	0,14
-0,3	-0,26
0,1	-0,06
0,6	0,14

$$\left(\mathrm{x}_{i1} - \mathrm{m}_1\right)\left(\mathrm{x}_{i1} - \mathrm{m}_1\right)^T$$

	-0,92	-0,14
-0,92	0,8464	0,1288
-0,14	0,1288	0,0196
	-0,32	-0,14
-0,32	0,1024	0,0448
-0,14	0,0448	0,0196
	0,68	-0,04
0,68	0,4624	-0,0272
-0,04	-0,0272	0,0016
	0,18	-0,04
0,18	0,0324	-0,0072
-0,04	-0,0072	0,0016
	0,38	0,36
0,38	0,1444	0,1368
0,36	0,1368	0,1296

$$n_1 \mathrm{S}_1 = \sum_{i=1}^{n_1} \left(\mathrm{x}_{i1} - \mathrm{m}_1\right)\left(\mathrm{x}_{i1} - \mathrm{m}_1\right)^T$$

1,588	0,276
0,276	0,172

$$n_2 \mathrm{S}_2$$

0,56	0,13
0,13	0,112

$$S_w = \frac{1}{n_1 + n_2 - 2}\left(n_1 S_1 + n_2 S_2\right)$$

0,2685	0,05075
0,05075	0,0355

$$S_w^{-1}$$

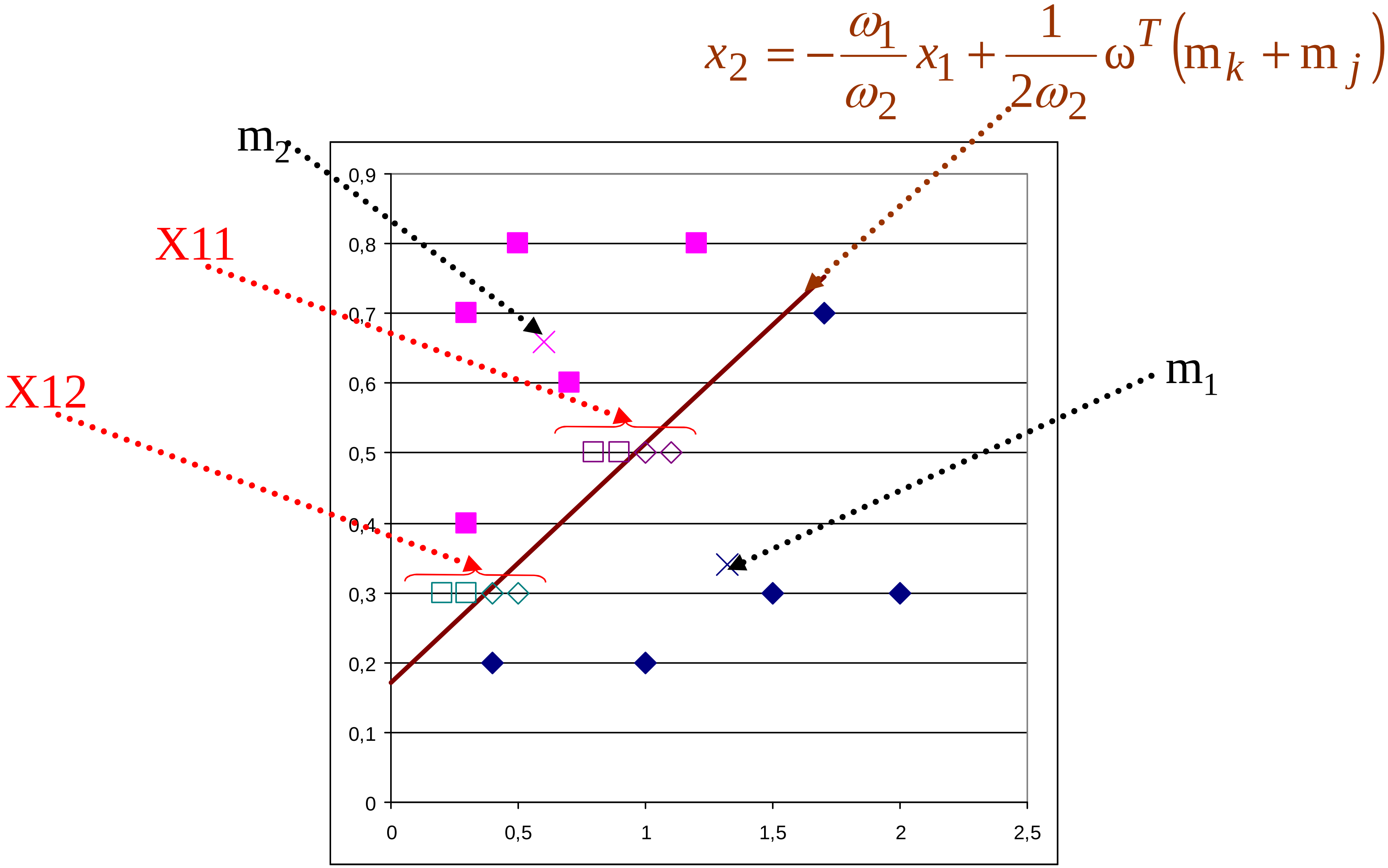
5,1034	-7,2957
-7,2957	38,5987

$$\omega = S_w^{-1}\left(m_1 - m_2\right)$$

6,009072
-17,6045

$$\left|\omega^T \mathrm{x} - \omega^T \mathrm{m}_k\right| < \left|\omega^T \mathrm{x} - \omega^T \mathrm{m}_j\right|, \quad k \neq j$$

	X(1)1	X(2)1	X(3)1	X(4)1	X(1)2	X(2)2	X(3)2	X(4)2
	0,8	0,9	1	1,1	0,2	0,3	0,4	0,5
	0,5	0,5	0,5	0,5	0,3	0,3	0,3	0,3
omegaTX	-3,99499	-3,39408	-2,79317	-2,19226	-4,07953	-3,47862	-2,87772	-2,27681
	m1	m1	m1	m1	m1	m1	m1	m1
	1,32	1,32	1,32	1,32	1,32	1,32	1,32	1,32
	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34
omegaTm1	1,946449	1,946449	1,946449	1,946449	1,946449	1,946449	1,946449	1,946449
	m2	m2	m2	m2	m2	m2	m2	m2
	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6
	0,66	0,66	0,66	0,66	0,66	0,66	0,66	0,66
omegaTm2	-8,01352	-8,01352	-8,01352	-8,01352	-8,01352	-8,01352	-8,01352	-8,01352
dist to cl1	5,941436	5,340528	4,739621	4,138714	6,025981	5,425074	4,824167	4,22326
dist to cl2	4,018532	4,61944	5,220347	5,821254	3,933987	4,534894	5,135801	5,736708



$$x_2 = -\frac{\omega_1}{\omega_2} x_1 + \frac{1}{2\omega_2} \omega^T \left(m_k + m_j \right)$$
m_2
X11
X12
m_1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0
0,5
1
1,5
2
2,5