Zaključni primjer
M = 3, nelinearne decizijske granice i
Učenje potencijalnim funkcijama

Uvod u raspoznavanje uzoraka

Prof. dr. sc. Slobodan Ribarić

ak.god. 2019/20

Example:

- M = 3 classes
- non-linearly separable classes
- find decision function by using generalized perceptron algorithm

$$U_4^1 = \{(0,0)^T, (0,4)^T, (4,0)^T, (4,4)^T\}$$

$$U_4^2 = \{(1,1)^T, (1,3)^T, (3,1)^T, (3,3)^T\}$$

$$U_1^3 = \{(2,2)^T\}$$

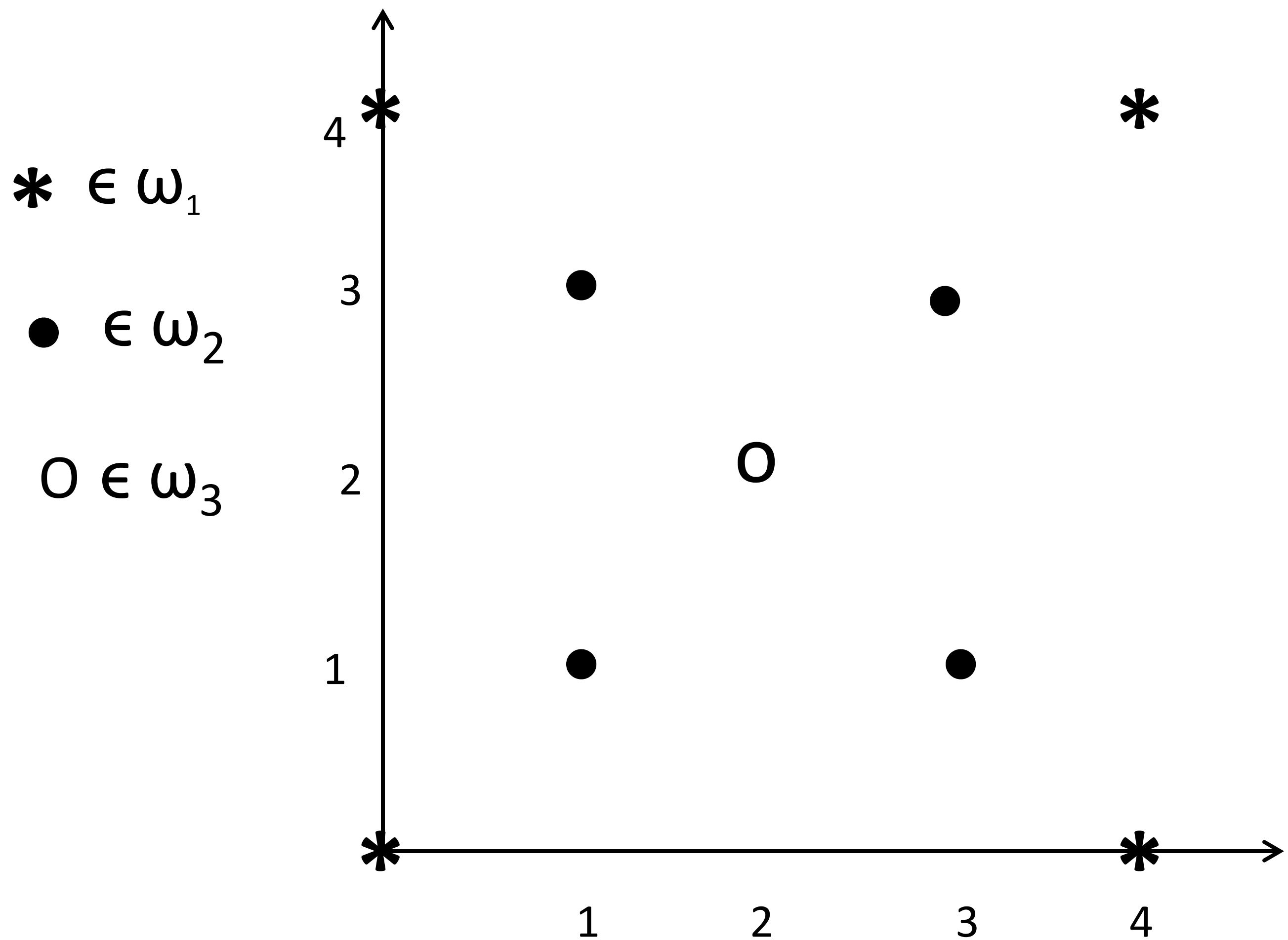
U_i^m m - index of class

i - denotes number of training vectors in a class

$$U_4^1 = \{(0,0)^T, (0,4)^T, (4,0)^T, (4,4)^T\}$$

$$U_4^2 = \{(1,1)^T, (1,3)^T, (3,1)^T, (3,3)^T\}$$

$$U_1^3 = \{(2,2)^T\}$$



Classes are not linearly separable!

Let us try with polynomial decision functions *r* = 2;

r = 2 and *n* = 2

$$g^2(\mathbf{x}) = \left(\sum_{p_1=1}^{2} \sum_{p_2=p_1}^{2} w_{p_1 p_2} \, x_{p_1} x_{p_2} \right) + g^1(\mathbf{x})$$

$$= w_{11}x_1x_1 + w_{12}x_1x_2 + w_{22}x_2x_2 + w_1x_1 + w_2x_2 + w_3$$

$$= w_{11}x_1^2 + w_{12}x_1x_2 + w_{22}x_2^2 + w_1x_1 + w_2x_2 + w_3$$

$$g^1(\mathbf{x}) = \left(\sum_{p_1=1}^{2} w_{p_1} \, x_{p_1} \right) + g^0(\mathbf{x}) = w_1x_1 + w_2x_2 + w_3$$

$$g^0(\mathbf{x}) = w_3$$

$$\mathrm{g}(\mathbf{x}) = w_{11}x_1^2 + w_{12}x_1x_2 + w_{22}x_2^2 + w_1x_1 + w_2x_2 + w_3$$

$$g(\mathbf{x}) = w_{11}x_1^2 + w_{12}x_1x_2 + w_{22}x_2^2 + w_1x_1 + w_2x_2 + w_3$$

$\mathbf{x} \rightarrow \mathbf{y}$

- in original two-dimensional space:

$$U_4^1 = \{(0,0)^T,(0,4)^T,(4,0)^T,(4,4)^T\}$$

$$U_4^2 = \{(1,1)^T,(1,3)^T,(3,1)^T,(3,3)^T\}$$

$$U_1^3 = \{(2,2)^T\}$$

- in transformed six-dimensional space:

$$U_4^1 = \{(0,0,0,0,0,1)^T,(0,0,16,0,4,1)^T,(16,0,0,4,0,1)^T,(16,16,16,4,4,1)^T\}$$

$$U_4^2 = \{(1,1,1,1,1,1)^T,(1,3,9,1,3,1)^T,(9,3,1,3,1,1)^T,(9,9,9,3,3,1)^T\}$$

$$U_1^3 = \{(4,4,4,2,2,1)^T\}$$

- let us apply the generalized parceptron algorithm (M = 3)

c = 1 and $\mathbf{w}_1 (1) = (1, 1, 1, 1, 1, 1)^T$, $\mathbf{w}_2(1) = (-1, 1, 2, -1, 1, 1)^T$, $\mathbf{w}_2(1) = (0, 0, 0, 0, 0, 1)^T$

1. Step

$\mathbf{y}_1 = (0, 0, 0, 0, 0, 1) \in \omega_1$

$g_1 (\mathbf{y}_1) = \mathbf{w}_1^T \mathbf{y}_1 = (1, 1, 1, 1, 1, 1) (0, 0, 0, 0, 0, 1)^T = 1$

$g_2(\mathbf{y}_1) = \mathbf{w}_2^T \mathbf{y}_1 = (-1, 1, 2, -1, 1, 1) (0, 0, 0, 0, 0, 1)^T = 1$

$g_3(\mathbf{y}_1) = \mathbf{w}_3^T \mathbf{y}_1 = (0, 0, 0, 0, 0, 1) (0, 0, 0, 0, 0, 1)^T = 1$

$\mathbf{w}_1 (2) = \mathbf{w}_1 (1) + \mathbf{y}_1 = (1, 1, 1, 1, 1, 2)^T$

$\mathbf{w}_2 (2) = \mathbf{w}_2 (1) - \mathbf{y}_1 = (-1, 1, 2, -1, 1, 0)^T$

$\mathbf{w}_3 (2) = \mathbf{w}_2 (1) - \mathbf{y}_1 = (0, 0, 0, 0, 0, 0)^T$

2. Step

$\mathbf{y}_2 = (0, 0, 16, 0, 4, 1) \in \omega_1$

2. Step

$\mathbf{y}_2 = (0, 0, 16, 0, 4, 1) \in \omega_1$

$g_1(\mathbf{y}_2) = \mathbf{w_2}^T\, \mathbf{y}_2 = (1, 1, 1, 1, 1, 2)\,(0, 0, 16, 0, 4, 1)^T = 22$

$g_2(\mathbf{y}_2) = \mathbf{w}_2^T\, \mathbf{y}_1 = (-1, 1, 2, -1, 1, 0)\,(0, 0, 16, 0, 4, 1)^T = 36$

$g_3(\mathbf{y}_2) = \mathbf{w}_3^T\, \mathbf{y}_1 = (0, 0, 0, 0, 0, 0)\,(0, 0, 16, 0, 4, 1)^T = 0$

$\mathbf{w}_1(3) = \mathbf{w}_1(2) + \mathbf{y}_2 = (1, 1, 1, 1, 1, 2)^T + (0, 0, 16, 0, 4, 1)^T =$

$= (1, 1, 17, 1, 5, 3)^T$

$\mathbf{w}_2(3) = \mathbf{w}_2(2) - \mathbf{y}_2 = (-1, 1, 2, -1, 1, 0)^T - (0, 0, 16, 0, 4, 1)^T =$

$= (-1, 1, -14, -1, -3, -1)^T$

$\mathbf{w}_3(3) = \mathbf{w}_3(2)$

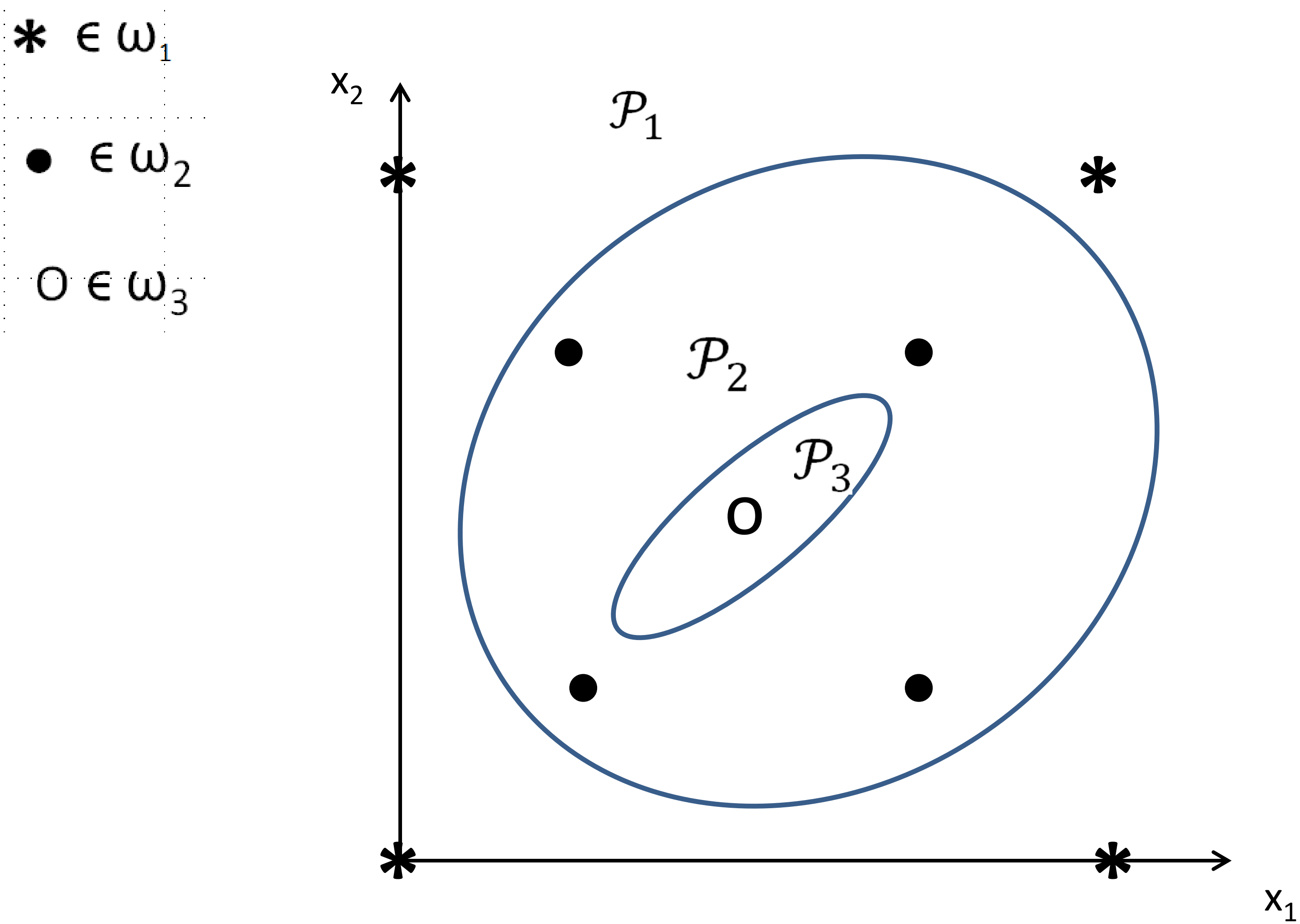
3. Step

…

4. Step

…

- After more then 2000 iteration steps the result is obtained



An alternative approach – the potential function

- properties of a potential function $K(\mathbf{x}, \mathbf{x}_k)$:

1. symmetry condition

$$K(\mathbf{x}, \mathbf{x}_k) = K(\mathbf{x}_k, \mathbf{x})$$

2. $K(\mathbf{x}, \mathbf{x}_k)$ has maximum value for $\mathbf{x} = \mathbf{x}_k$

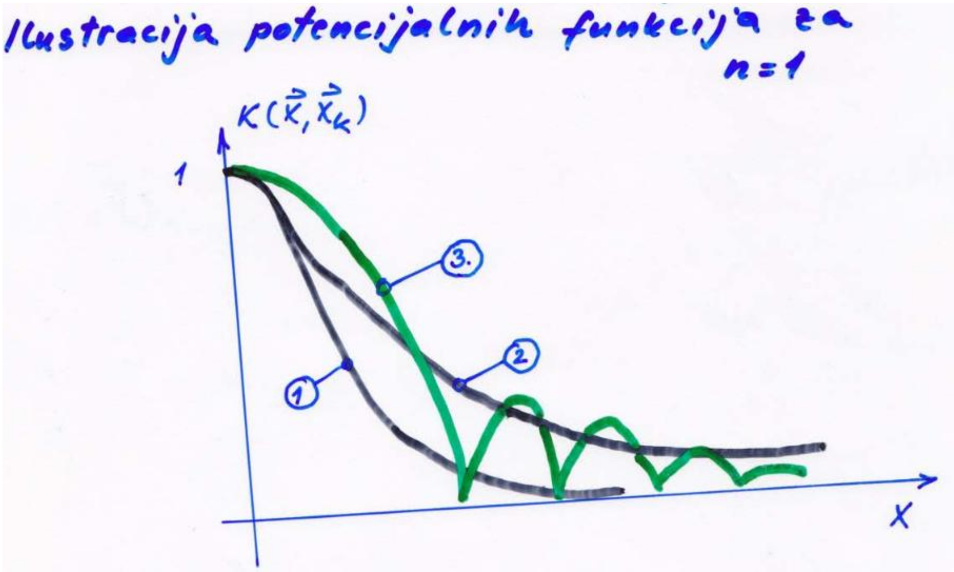
3. $K(\mathbf{x}, \mathbf{x}_k) \rightarrow 0$ if $\|\mathbf{x} - \mathbf{x}_k\| \rightarrow \infty$

- examples of the potential functions

$$K(\mathbf{x},\mathbf{x}_k) = \exp\{-\alpha \,\|\mathbf{x} - \mathbf{x}_k\|^2\}$$

$$K(\mathbf{x},\mathbf{x}_k) = \frac{1}{1 + \alpha \,\|\mathbf{x} - \mathbf{x}_k\|^2}$$

$$K(\mathbf{x},\mathbf{x}_k) = \left|\frac{\sin \alpha\|\mathbf{x} - \mathbf{x}_k\|^2}{\alpha\|\mathbf{x} - \mathbf{x}_k\|^2}\right|$$



- basic idea of using potential functions in design of decision function for M = 2 classes

- training vectors are points in *n* dimensional feature space
- let us assume that in each such point of space is located a positive unit charge +q if $\mathbf{x} \in \omega_1$, and a negative unit charge -q if $\mathbf{x} \in \omega_2$
- both types of charges form a potential field – potential at any of these points attains a peak value and then decreases at any point in the space away from the pattern point $\mathbf{x}$
  - the cluster of vectors $\mathbf{x}_i \in \omega_1$, $i = 1, 2, \dots N_1$; where N_1 is a number of training vectors from ω_1, form *positive* "potential plateau";
  - the cluster of vectors $\mathbf{x}_i \in \omega_2$, $i = 1, 2, \dots N_2$; where N_2 is a number of training vectors from ω_2, form *negative* "potential plateau";
- these two plateaus are separated by "valley" in which the value of potential field is zero (due to interactions of positive and negative charges)

Learning decision function for M = 2 classes

- there are N training feature vectors; $\mathbf{x}_i \in \omega_1 \cup \omega_2$; $i = 1, 2, \ldots, N_i$

- let us suppose that a decision function has a following form:

$$d_k(\mathbf{x}) = d_{k-1}(\mathbf{x}) + r_k\, K(\mathbf{x}, \mathbf{x}_k)$$

where

$$d_0(\mathbf{x}) = 0$$

and r_k is a correction factor

$$r_k = \begin{cases} 0 \text{ for je } \mathbf{x}_k \in \omega_1 \text{ and } d_{k-1}(\mathbf{x}_k) > 0 \\ 0 \text{ for je } \mathbf{x}_k \in \omega_2 \text{ and } d_{k-1}(\boldsymbol{x}_k) < 0 \\ 1 \text{ for je } \mathbf{x}_k \in \omega_1 \text{ and } d_{k-1}(\mathbf{x}_k) \leq 0 \\ -1 \text{ for je } \mathbf{x}_k \in \omega_2 \text{ and } d_{k-1}(\mathbf{x}_k) \geq 0 \end{cases}$$

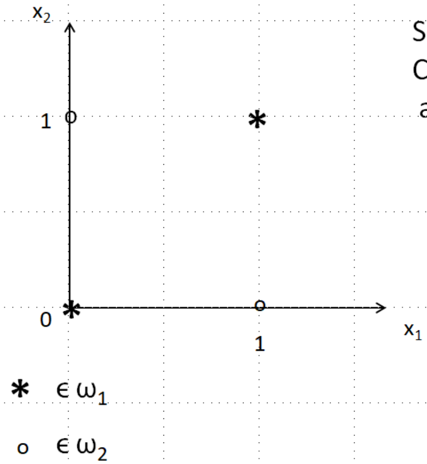
- for $k = 1$

$$d_1(\mathbf{x}_1) = 0 + K(\mathbf{x}, \mathbf{x}_1) \quad \text{if } \mathbf{x}_1 \in \omega_1$$

$$d_1(\mathbf{x}_1) = 0 - K(\mathbf{x}, \mathbf{x}_1) \quad \text{if } \mathbf{x}_1 \in \omega_2$$

Example:

-Training set: $\omega_1 = \{(0, 0)^T, (1, 1)^T\}$
$\omega_2 = \{(0, 1)^T, (1, 0)^T\}$



- let us select potential function:

$$K(\mathbf{x}, \mathbf{x}_k) = \exp\{-\alpha \, \|\mathbf{x} - \mathbf{x}_k\|^2\}$$

and $\alpha = 1.0$

- for n = 2 (dimensionality of feature vectors) the potential function has a form $K(\mathbf{x},\mathbf{x}_k) = e^{-\alpha\|\mathbf{x}-\mathbf{x}_k\|^2} = e^{-\|\mathbf{x}-\mathbf{x}_k\|^2} = e^{-[(x_1-x_{k1})^2+(x_2-x_{k2})^2}$

$\mathbf{x}_1 = (0, 0)^T \in \omega_1$ - first training vector

$d_1(\mathbf{x}_1) = d_0(\mathbf{x}_1) + K(\mathbf{x}, \mathbf{x}_1)$; $d_0(\mathbf{x}_1) = 0$

$$d_1(\mathbf{x}_1) = e^{-[(x_1-0)^2+(x_2-0)^2]} = e^{-({x_1}^2+x_2^2)} > 0$$

$\mathbf{x}_2 = (1, 1)^T \in \omega_1$

$$d_1(\mathbf{x}_2) = e^{-({x_1}^2+x_2^2)} = e^{-(1+1)} = e^{-2} > 0 \ ; r_2 = 0$$

$$d_2(\mathbf{x}) = d_1(\mathbf{x}) = e^{-({x_1}^2+x_2^2)}$$

$\mathbf{x}_3 = (0, 1)^T \in \omega_2$

$$d_2(\mathbf{x}_3) = e^{-({x_1}^2+x_2^2)} = e^{-(0+1)} = e^{-1} > 0$$; $r_3 = -1$

$$d_3(\mathbf{x}) = d_2(\mathbf{x}) - K(\mathbf{x}, \mathbf{x}_3) = e^{-({x_1}^2+x_2^2)} - e^{-[(x_1-0)^2+(x_2-1)^2]}$$

$\mathbf{x}_4 = (1,0)^T \in \omega_2$

$$d_3(\mathbf{x}_4) = e^{-({x_1}^2+x_2^2)} - e^{-[(x_1-0)^2+(x_2-1)^2]} = e^{-1} - e^{-2} > 0$$

$r_4 = -1$

$$d_4(\mathbf{x}) = d_3(\mathbf{x}) - K(\mathbf{x}, \mathbf{x}_4) = e^{-({x_1}^2+x_2^2)} - e^{-[(x_1-0)^2+(x_2-1)^2]} - e^{-[(x_1-1)^2+x_2^2]}$$

$\mathbf{x}_5 = \mathbf{x}_1 = (0, 0)^T \in \omega_1$

$$d_4(\mathbf{x}_5) = e^{-0} - e^{-1} - e^{-1} > 0$$

$$d_5(\mathbf{x}) = d_4(\boldsymbol{x})$$

$\mathbf{x}_6 = \mathbf{x}_2 = (1, 1)^T \in \omega_1$

$$\mathrm{d}_5(\mathbf{x}_6) = e^{-2} - e^{-1} - e^{-1} < 0$$

$$\mathrm{d}_6(\mathbf{x}) = \mathrm{d}_5(\mathbf{x}) + K(\mathbf{x}, \mathbf{x}_6)$$

$$\mathrm{d}_6(\mathbf{x}) = e^{-({x_1}^2 + x_2^2)} - e^{-[(x_1-0)^2 + (x_2-1)^2]} - e^{-[(x_1-1)^2 + x_2^2]} + e^{-[(x_1-1)^2 + (x_2-1)^2]}$$

$\mathbf{x}_7 = \mathbf{x}_3 = (0, 1)^T \in \omega_2$

$$\mathrm{d}_6(\mathbf{x}_7) = e^{-1} - e^{-0} - e^{-2} - e^{-1} < 0$$

$$\mathrm{d}_7(\mathbf{x}) = \mathrm{d}_6(\mathbf{x})$$

$\mathbf{x}_8 = \mathbf{x}_4 = (1, 0)^T \in \omega_2$

$$\mathrm{d}_7(\mathbf{x}_8) = e^{-1} - e^{-1} - e^{-0} - e^{-1} < 0$$

$$\mathrm{d}_8(\mathbf{x}) = \mathrm{d}_7(\mathbf{x})$$

$\mathbf{x}_9 = \mathbf{x}_1 = (0,0)^T \in \omega_1$

$$\mathrm{d}_8(\mathbf{x}_9) = e^{-0} - e^{-1} - e^{-1} - e^{-2} > 0$$

$$\mathrm{d}_9(\mathbf{x}) = \mathrm{d}_8(\mathbf{x})$$

$\mathbf{x}_{10} = \mathbf{x}_2 = (1,1)^T \in \omega_1$

$$\mathrm{d}_9(\mathbf{x}_{10}) = e^{-2} - e^{-1} - e^{-1} - e^{-0} > 0$$

$$\mathrm{d}_{10}(\mathbf{x}) = \mathrm{d}_9(\mathbf{x})$$

- an entire iteration is performed without an error –
the decision function is:

$$\mathrm{d}(\mathbf{x}) = e^{-({x_1}^2 + x_2^2)} - e^{-[{x_1}^2 + (x_2-1)^2]} - e^{-[(x_1-1)^2 + x_2^2]} + e^{-[(x_1-1)^2 + (x_2-1)^2]}$$

Potential functions - Multiclass generalization

M > 2

- at the beginning of training phase all decision functions are assumed to be zero:

$$\mathrm{d}_0^1(\mathbf{x}) = 0;\ \mathrm{d}_0^2(\mathbf{x}) = 0; \dots ;\ \mathrm{d}_0^M(\mathbf{x}) = 0$$

- suppose that at *k*th iterative step $\mathbf{x}_k$ belongs to class ω_i

if

$$\mathrm{d}_{k-1}^i(\mathbf{x}_k) > \mathrm{d}_{k-1}^j(\mathbf{x}_k) \text{ for } \forall j \neq i$$

then

$$d_k^i(\mathbf{x}) = d_{k-1}^i(\mathbf{x}), i = 1, 2, \dots, M$$

if $\mathbf{x}_k \in \omega_i$ and for some l

if $\mathbf{x}_k \in \omega_i$ and for some l

$$\mathrm{d}_{k-1}^{i}(\mathbf{x}_k) < \mathrm{d}_{k-1}^{l}(\mathbf{x}_\mathrm{k}),$$

the following corrections are made:

$$\mathrm{d}_{k}^{i}(\mathbf{x}) = \mathrm{d}_{k-1}^{i}(\mathbf{x}) + K(\mathbf{x}, \mathbf{x}_k)$$

$$\mathrm{d}_{k}^{l}(\mathbf{x}) = \mathrm{d}_{k-1}^{l}(\mathbf{x}) - K(\mathbf{x}, \mathbf{x}_k)$$

$$\mathrm{d}_{k}^{j}(\mathbf{x}) = \mathrm{d}_{k-1}^{j}(\mathbf{x}), \quad j = 1,2,3,\dots,M; \quad j \neq i; j \neq l$$