# ANALÝZA A KLASIFIKACE DAT

prof. Ing. Jiří Holčík, CSc.

MU

IBA

EVROPSKÁ UNIE esf

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání pro konkurenceschopnost 2007-13

UNIVERSITAS MASARYKIANA BRUNENSIS

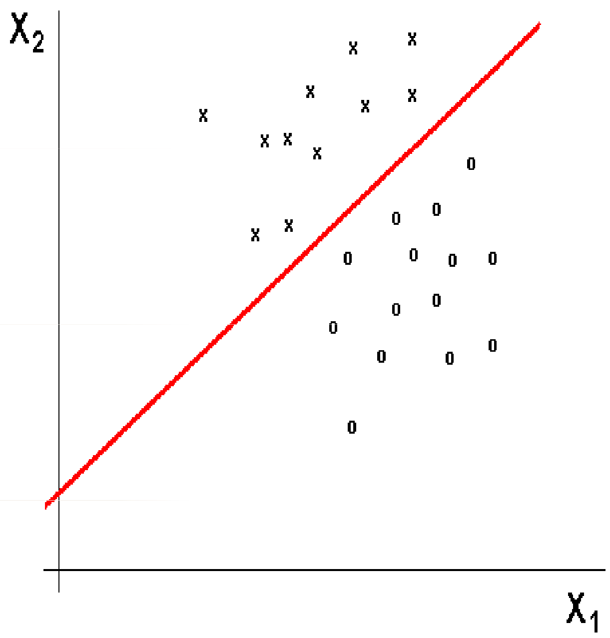
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# V. LINEÁRNÍ KLASIFIKACE

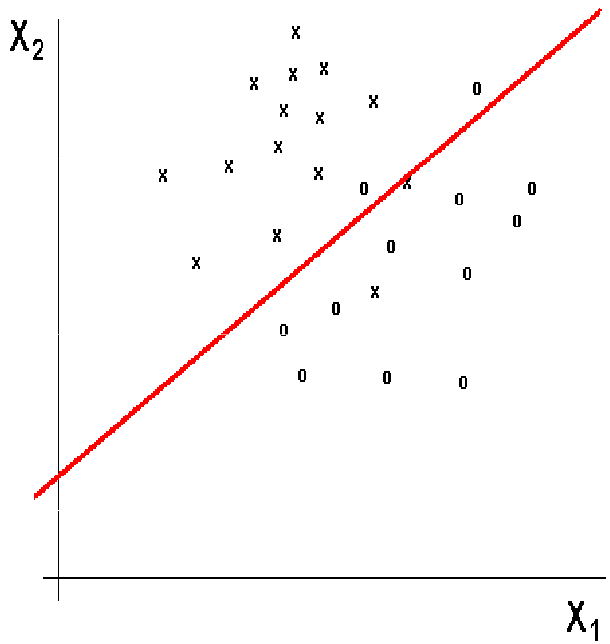
# PRINCIPY KLASIFIKACE

- ☑ pomocí diskriminačních funkcí – funkcí, které určují míru příslušnosti k dané klasifikační třídě;
- ☑ pomocí definice hranic mezi jednotlivými třídami a logických pravidel;
- ☑ pomocí vzdálenosti od reprezentativních obrazů (etalonů) klasifikačních tříd;
- ☑ pomocí ztotožnění s etalony;

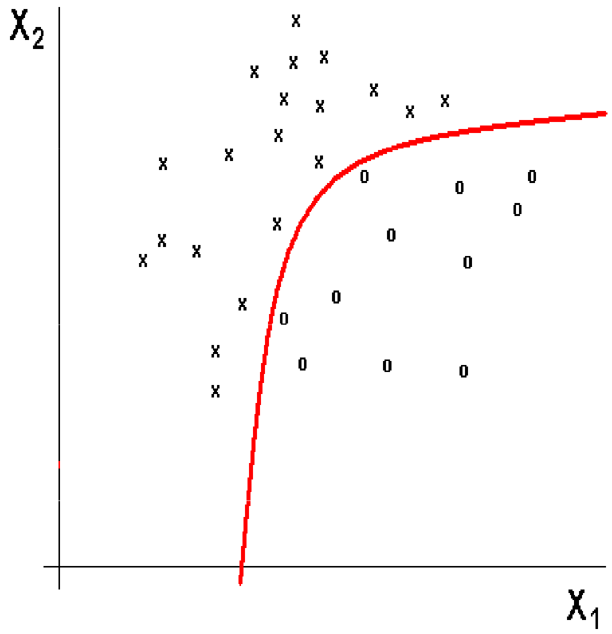
# LINEÁRNÍ SEPARABILITA




lineárně separabilní úloha




lineárně neseparabilní úloha
lineárně separované klasifikační třídy




nelineárně separabilní úloha

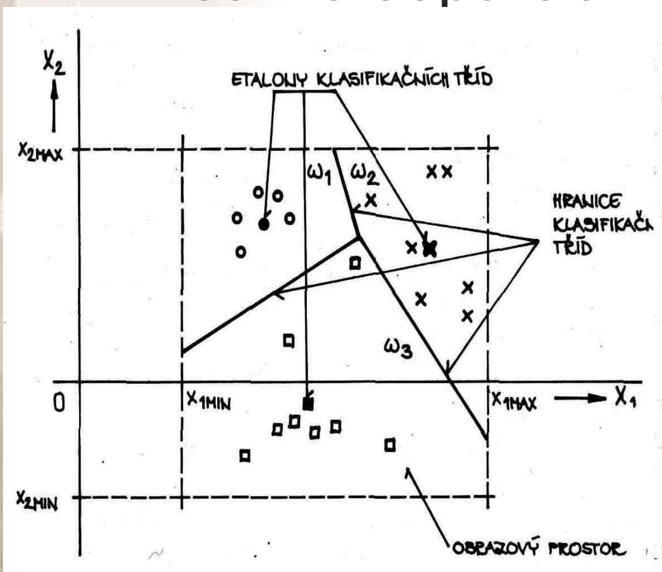
# KLASIFIKACE PODLE DISKRIMINAČNÍCH FUNKCÍ

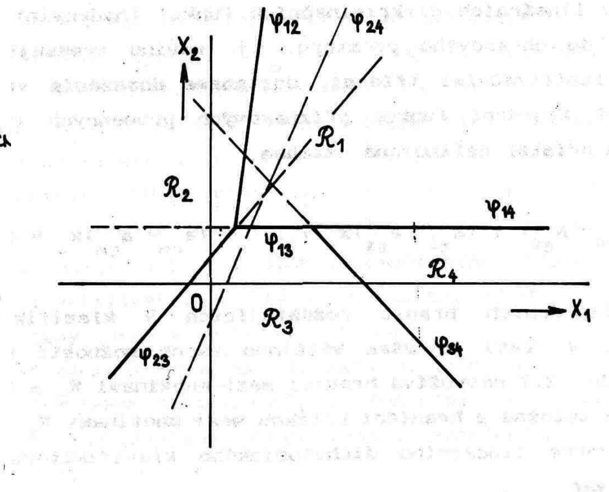
- nejjednodušším tvarem diskriminační funkce je funkce lineární, která má tvar

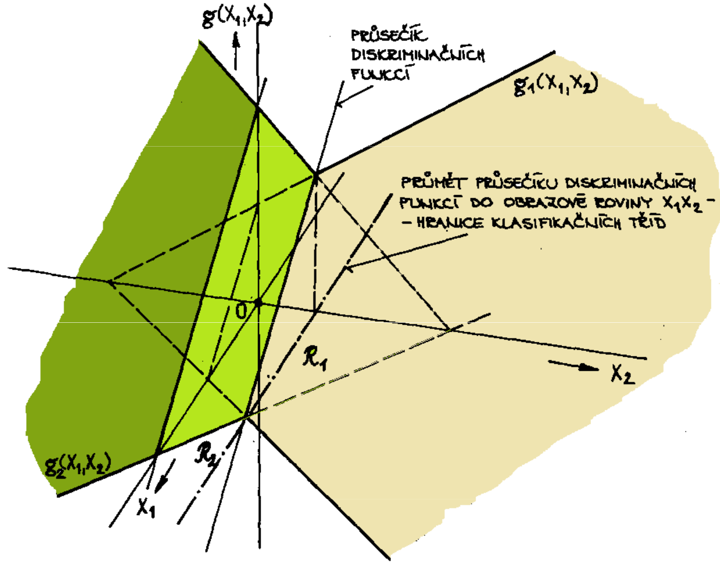
$$g_r(\mathbf{x}) = a_{r0} + a_{r1}x_1 + a_{r2}x_2 + \ldots + a_{rn}x_n$$

kde $a_{r0}$ je práh diskriminační funkce posouvající počátek souřadného systému a $a_{ri}$ jsou váhové koeficienty i-tého příznaku $x_i$

- lineárně separabilní třídy



# DICHOTOMICKÁ ÚLOHA
## PRINCIP

nejjednodušší realizace hraniční plochy je lineární funkcí

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

**w** je váhový vektor, $w_0$ je práh;

$$\mathbf{x} \in \omega_1, \text{ když } y(\mathbf{x}) \geq 0$$

$$\mathbf{x} \in \omega_2, \text{ když } y(\mathbf{x}) < 0$$

rovnice hranice je $y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = 0$

((n-1)-rozměrná nadplocha (nadrovina) v n-rozměrném prostoru

# DICHOTOMICKÁ ÚLOHA
## ZÁKLADNÍ VLASTNOSTI

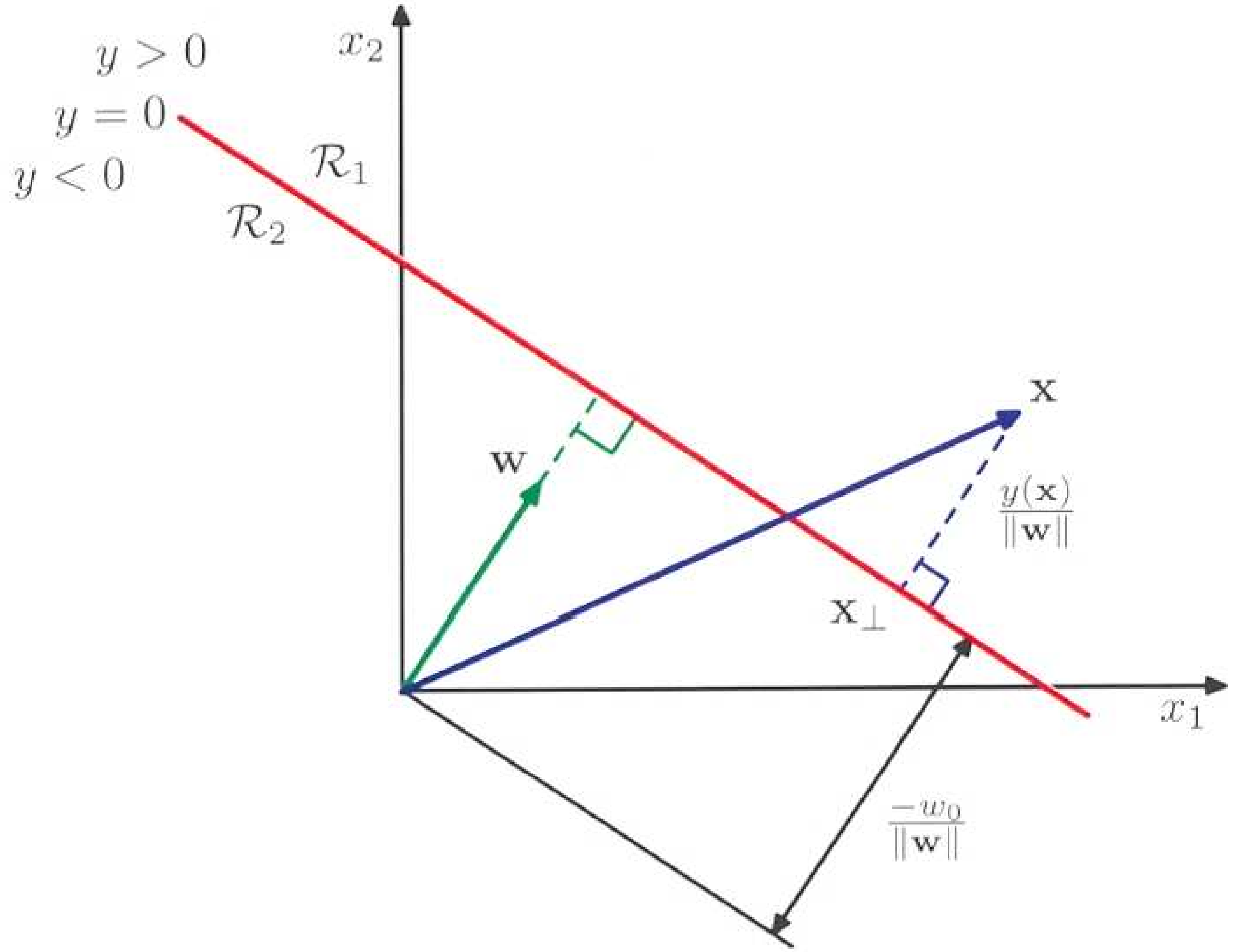
zápis v jiném (kompaktnějším) tvaru:

$x_0 = 1$ a pak $\tilde{\mathbf{w}} = (w_0, \mathbf{w})$ a $\tilde{\mathbf{x}} = (x_0, \mathbf{x})$

z toho

$$y(\mathbf{x}) = \tilde{\mathbf{w}}^T.\tilde{\mathbf{x}}$$

# DICHOTOMICKÁ ÚLOHA
## ZÁKLADNÍ VLASTNOSTI



# DICHOTOMICKÁ ÚLOHA
## ZÁKLADNÍ VLASTNOSTI

- ☑ pro $\mathbf{x}_A$, $\mathbf{x}_B$ na hraniční přímce je $y(\mathbf{x}_A)= y(\mathbf{x}_B)= 0$; proto je i $\mathbf{w}^T(\mathbf{x}_A-\mathbf{x}_B)=0 \Rightarrow$ vektor $\mathbf{w}$ je ortogonální (kolmý) k hraniční přímce;
- ☑ je-li $\mathbf{x}$ na hraniční přímce, je $y(\mathbf{x})= 0$ a tak normálová vzdálenost počátku od hraniční přímky je dána vztahem

$$\frac{\mathbf{w}^T\mathbf{x}}{\|\mathbf{w}\|} = -\frac{w_0}{\|\mathbf{w}\|}$$

$y(\mathbf{x})$ udává kolmou vzdálenost d bodu $\mathbf{x}$ od hraniční přímky (je-li $\mathbf{x}_\perp$ ortogonální projekce $\mathbf{x}$ na hranici tak, že

$$\mathbf{x} = \mathbf{x}_\perp + d\frac{\mathbf{w}}{\|\mathbf{w}\|}$$

vynásobením obou stran $\mathbf{w}^T$, přičtením $w_0$ a s použitím $y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0$ a $y(\mathbf{x}_\perp) = \mathbf{w}^T\mathbf{x}_\perp + w_0 = 0$, dostaneme

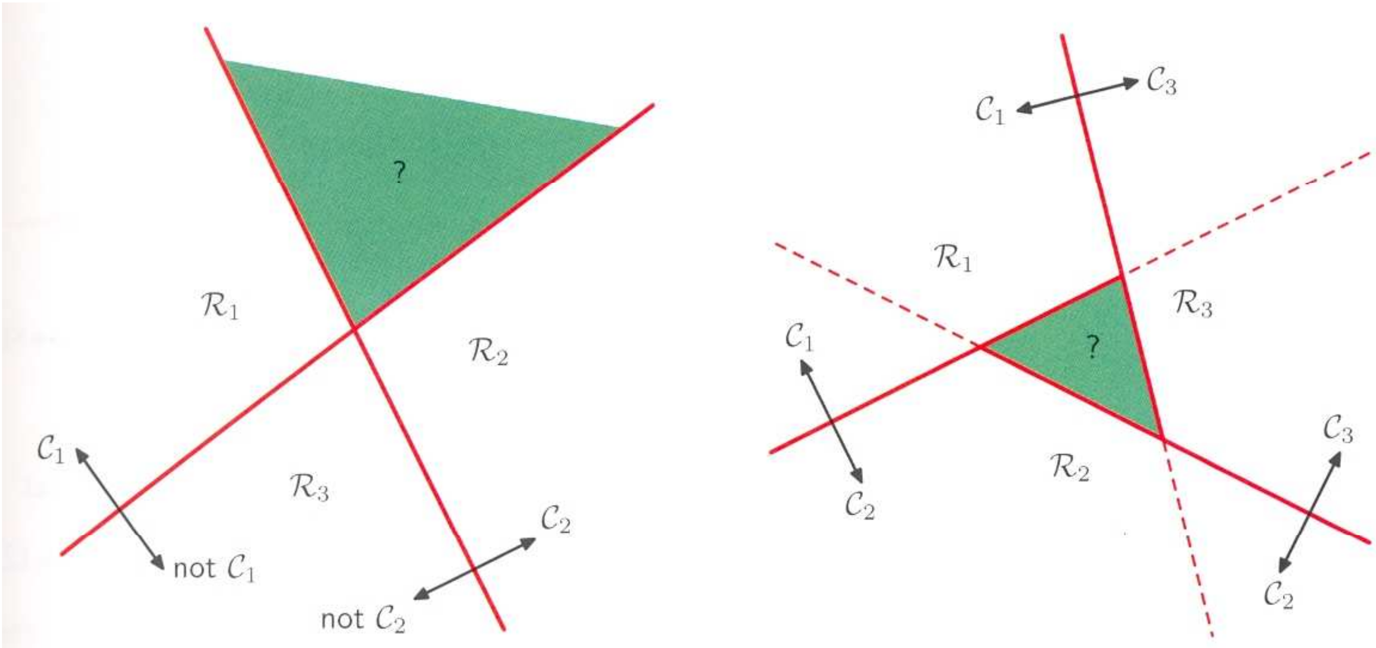
$$d = \frac{y(\mathbf{x})}{\|\mathbf{w}\|}$$

# ÚLOHA S VÍCE TŘÍDAMI

- kombinace více tříd (problém?):
  - klasifikace „jedna versus zbytek“

    R-1 hranice oddělí jednu klasifikační třídu od všech dalších
  - klasifikace „jedna versus jedna“

    R(R-1)/2 binárních hranic mezi každými dvěma třídami



# ÚLOHA S VÍCE TŘÍDAMI

- jak se vyhnout „problémům“?

zavedením principu diskriminační funkce

$$g_r(\mathbf{x}) = \mathbf{w}_r^T\mathbf{x} + w_{r0}$$

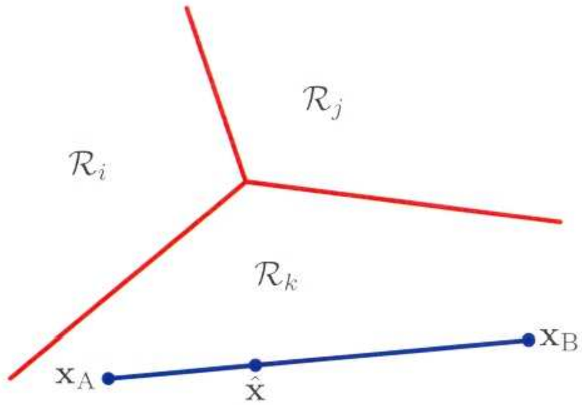
do r-té třídy $\omega_r$ zařadíme obraz $\mathbf{x}$ za předpokladu, že

$$g_r(\mathbf{x}) > g_s(\mathbf{x}) \text{ pro } \forall\ r \neq s$$

klasifikační hranice je průmět průsečíku

$$g_r(\mathbf{x}) = g_s(\mathbf{x}) \text{ do obrazového prostoru}$$

takto definovaný klasifikační prostor je vždy spojitý a konvexní



# METODY STANOVENÍ KLASIFIKAČNÍCH HRANIC

- ☑ metoda nejmenších čtverců
- ☑ perceptron (neuron)
- ☑ Fisherova lineární diskriminace
- ☑ algoritmus podpůrných vektorů

# METODA NEJMENŠÍCH ČTVERCŮ

- minimalizace součtu čtverců chybové funkce;
- mějme cílový (klasifikační) vektor vyjádřen binárním kódem 1 z R ($\mathbf{t} = (0,0,0,1,0)^T$)
- každá je třída $\omega_r$ popsána lineární funkcí

$$g_r(\mathbf{x}) = \mathbf{w}_r^T\mathbf{x} + w_{r0},$$

kde $r = 1, \ldots, R$;

# METODA NEJMENŠÍCH ČTVERCŮ

sumární popis těchto reprezentací je

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

kde $\tilde{\mathbf{W}}^{T}$ je matice, jejíž r-tý sloupec zahrnuje n+1 dimenzionální vektor

$\tilde{\mathbf{w}}_r = (w_0, \mathbf{w}_r^T)$ a $\tilde{\mathbf{x}} = (x_0, \mathbf{x}^T)$ , $x_0 = 1$

hodnota x na vstupu je zařazena do třídy, pro níž je $g_r(\mathbf{x}) = \tilde{\mathbf{w}}_r^T.\tilde{\mathbf{x}}$ největší;

# METODA NEJMENŠÍCH ČTVERCŮ

pokud máme učební množinu vyjádřenou $\{\mathbf{x}_i, \mathbf{t}_i\}$, $i=1,\dots,n$ a i-tý řádek matice $\mathbf{T}$ obsahuje vektor $\mathbf{t}_i^T$ a v matici $\tilde{\mathbf{X}}$ je i-tý řádek $\tilde{\mathbf{x}}_i^T$, pak funkce součtu čtverců chyb je

$$E_n(\tilde{\mathbf{W}}) = \frac{1}{2}\mathrm{Tr}\left\{\left(\tilde{\mathbf{X}}.\tilde{\mathbf{W}} - \mathbf{T}\right)^T \left(\tilde{\mathbf{X}}.\tilde{\mathbf{W}} - \mathbf{T}\right)\right\}$$

Derivací podle $\tilde{\mathbf{W}}$, kterou položíme rovno nule dostáváme

$$\tilde{\mathbf{W}} = (\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T\mathbf{T} = \tilde{\mathbf{X}}^{\Im}\mathbf{T}$$

kde $\tilde{\mathbf{X}}^{\Im}$ je tzv. pseudoinverzní matice k matici $\tilde{\mathbf{X}}$. Diskriminační funkce pak jsou ve tvaru

$$\mathbf{g}(\mathbf{x}) = \tilde{\mathbf{W}}^T.\tilde{\mathbf{x}} = \mathbf{T}^T(\tilde{\mathbf{X}}^{\Im})^T.\tilde{\mathbf{x}}$$

# METODA NEJMENŠÍCH ČTVERCŮ

když cílové vektory trénovací množiny $\mathbf{t}_k$, k=1,...K splňují lineární funkcí

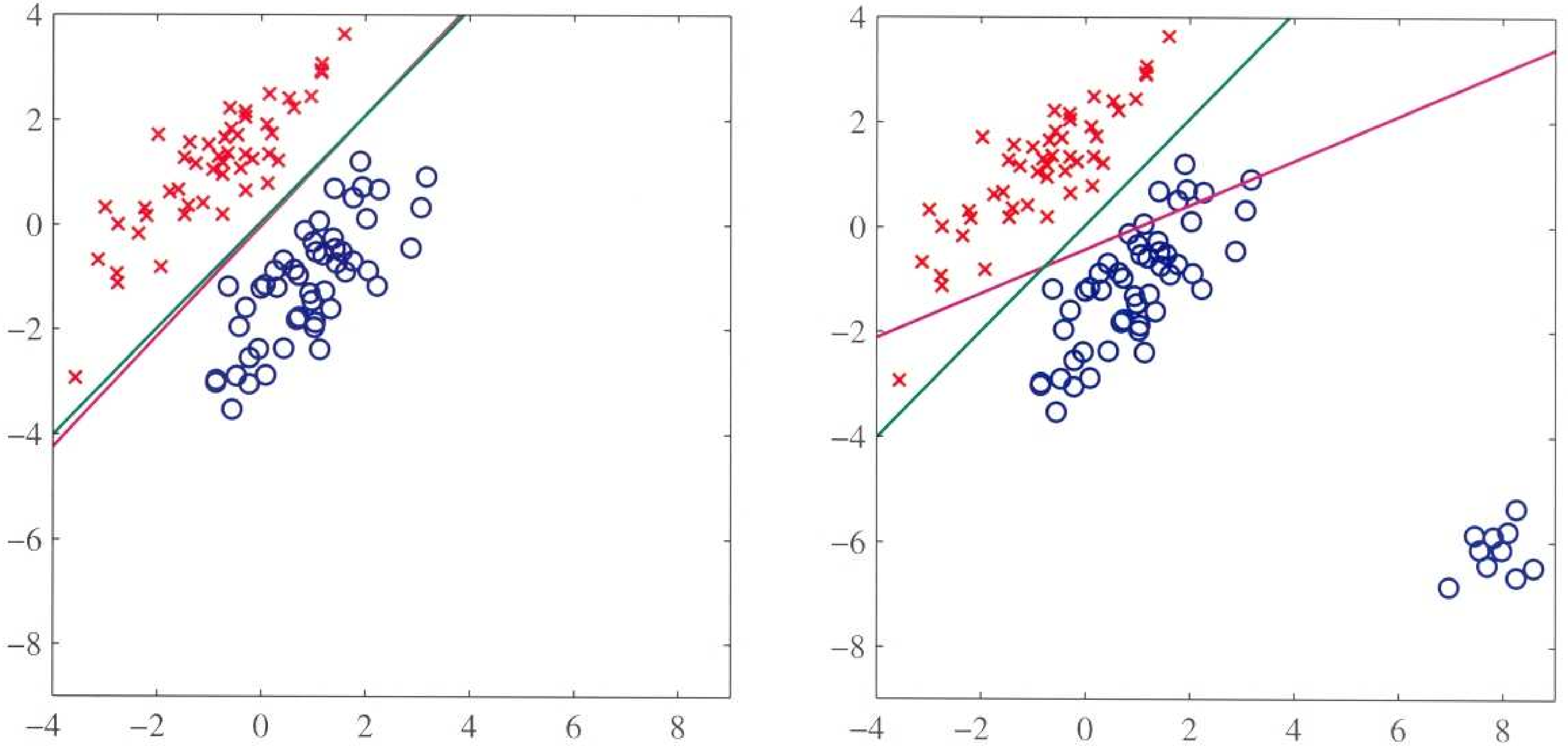
$$\mathbf{a}^T.\mathbf{t}_k + b = 0$$

pro libovolné k a dané konstanty **a** a b, pak lineární klasifikační model pro libovolný obraz splňuje ekvivalentní vztah

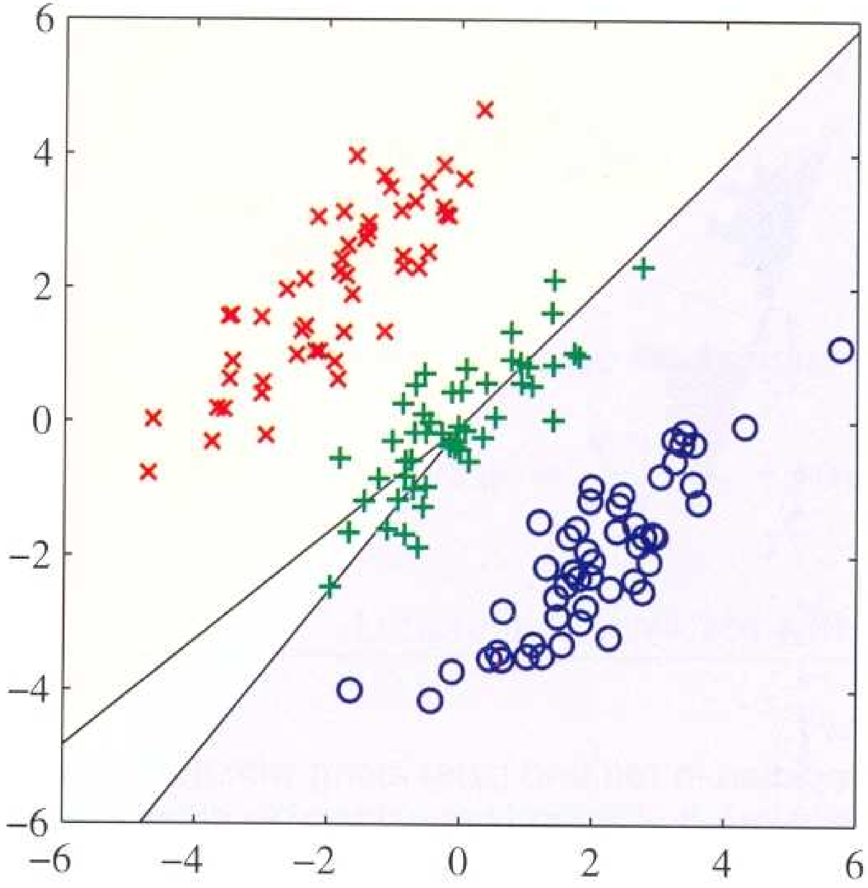
$$\mathbf{a}^T.\mathbf{g}(\mathbf{x}) + b = 0$$

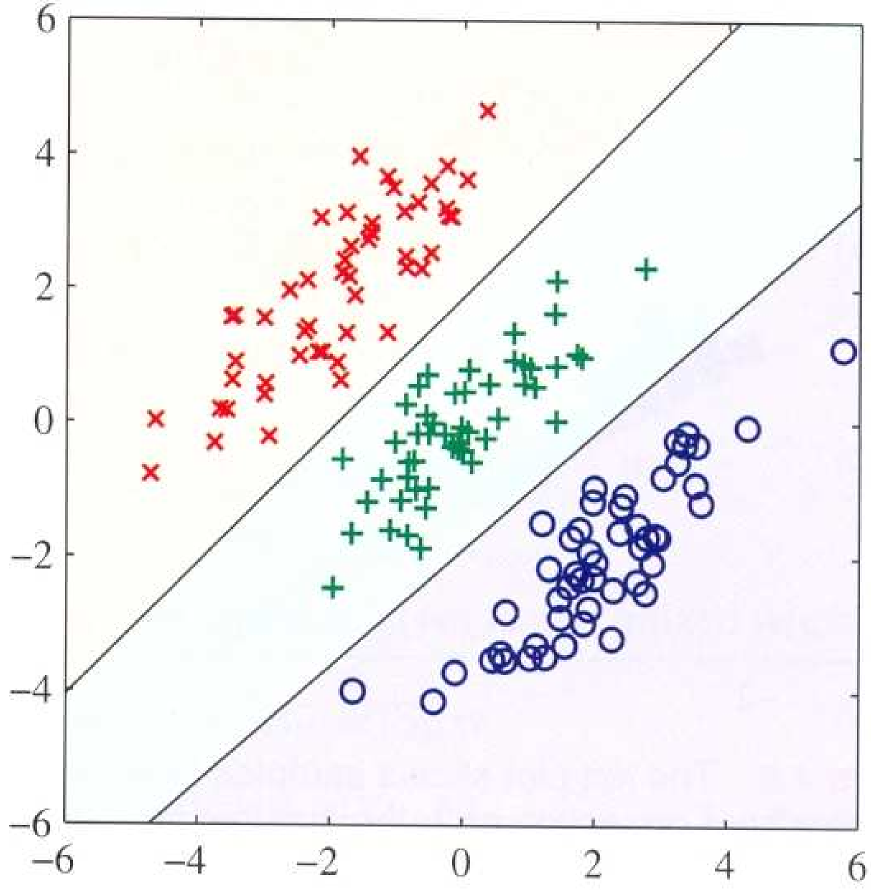
To znamená, že pro kódovací schéma 1 z R pro R klasifikačních tříd, je součet prvků vektoru $\mathbf{g}(\mathbf{x})$ roven jedné stejně jako součet prvků vektoru $\mathbf{t}_k$ pro libovolný obrazový vektor **x**. Tento požadavek ale není postačující, protože hodnoty vektoru $\mathbf{g}(\mathbf{x})$ nejsou nutně vázány na interval $\langle 0; 1 \rangle$, což by bylo třeba, kdyby měly vyjadřovat odhady pravděpodobností zatřídění do jednotlivých klasifikačních kategorií.

# METODA NEJMENŠÍCH ČTVERCŮ



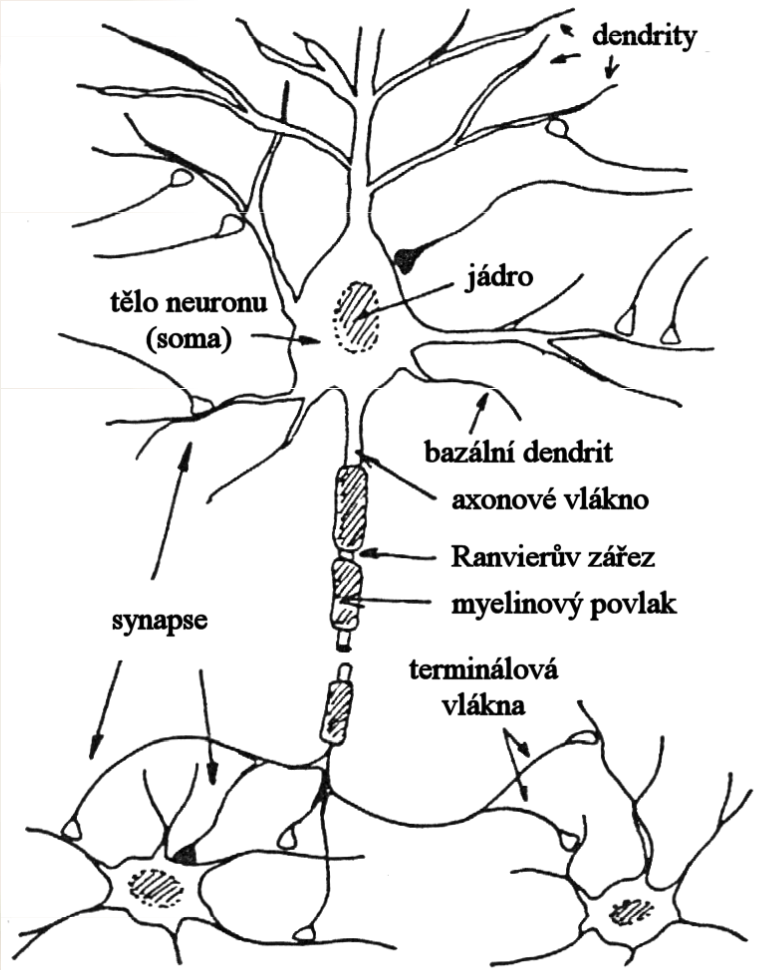
# METODA NEJMENŠÍCH ČTVERCŮ

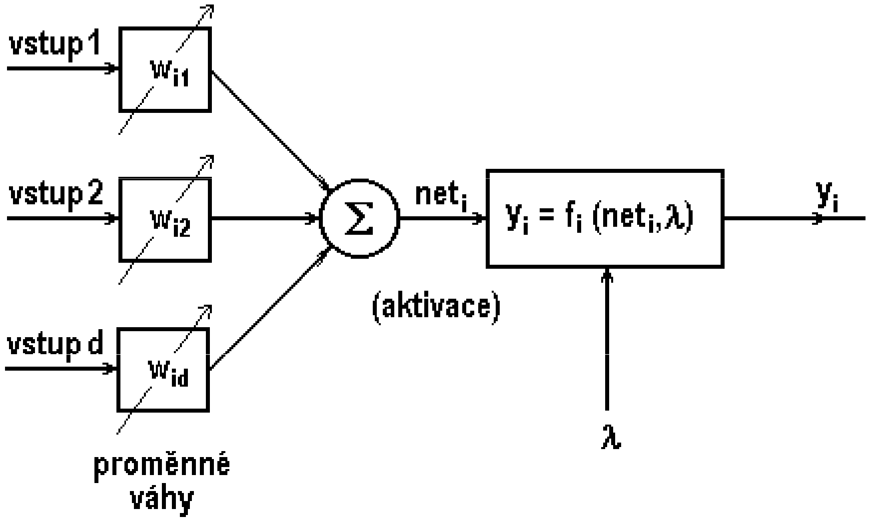


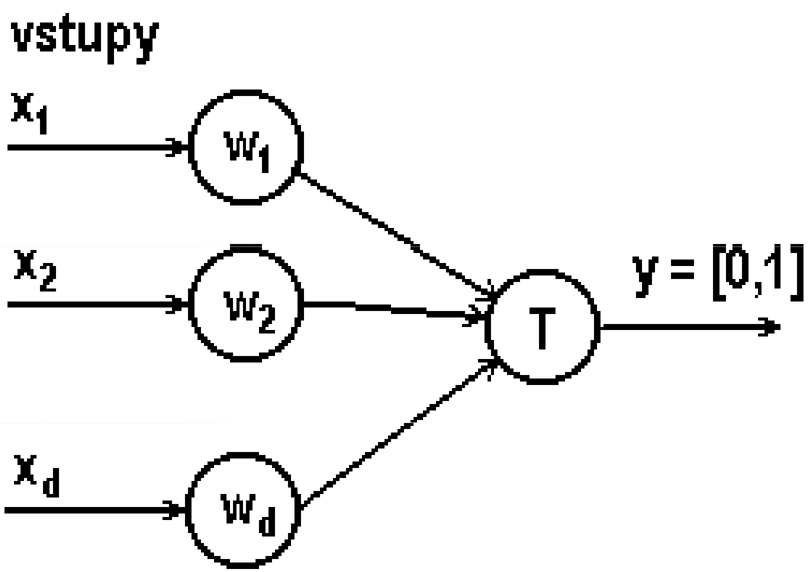
# PERCEPTRON

## MODEL NEURONU



# MODEL NEURONU




| vstup | výstup |
| --- | --- |
| $\sum_{k=1}^{d} x_i w_i < T$ | 0 |
| $\sum_{k=1}^{d} x_i w_i \geq T$ | 1 |

$$\xi = \sum_{i=1}^{n} w_i x_i - T = \sum_{i=0}^{n} w_i x_i$$

*Lineární model neuronu s prahem*

# PERCEPTRON

## UČENÍ

Obecné požadavky na postup nastavení hodnot váhových koeficientů perceptronu (a nejen perceptronu) jsou:

- ☑ algoritmická formulace – tj. metoda musí najít řešení pomocí konečného počtu dílčích kroků;
- ☑ konvergence – výpočet by měl být monotónní a pokud možno co nejrychlejší.

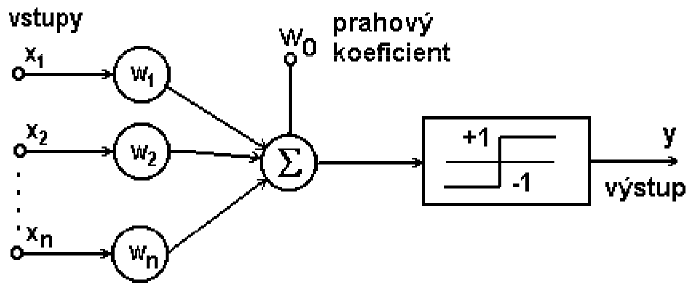
# PERCEPTRON

## UČENÍ

na vstup perceptronu jsou vkládány prvky trénovací množiny a výsledek klasifikace je srovnán s očekávanou správnou klasifikací;

- ☑ pokud je rozdíl mezi klasifikátorem určenou a správnou klasifikací větší než určitá předem daná mez definující přípustnou chybu, pak se parametry klasifikátoru (váhové koeficienty) změní tak, aby se chyba mezi určenou a požadovanou klasifikací minimalizovala;
- ☑ pokud je chyba klasifikace větší než předem stanovená mez, pak učení dále pokračuje, v opačném případě se učení ukončí a klasifikátor je možné použít ke klasifikaci. Kromě zmíněného absolutního kritéria ukončení učicí se fáze se v současné době často používá pro zastavení učení i relativní kritérium založené poklesu chyby během daného časového okna.

# PERCEPTRON




- ☑ předpokládejme, že

$\mathbf{w}^{*T}\mathbf{x} > 0$ pro $\forall x \in \omega_1$

$\mathbf{w}^{*T}\mathbf{x} < 0$ pro $\forall x \in \omega_2$

- ☑ snažíme se o nalezení extrému ztrátové funkce perceptronu

$$J(\mathbf{w}) = \sum_{\mathbf{x} \in Y} (\delta_x \mathbf{w}^T \mathbf{x}) \qquad (🚓)$$

- ☑ Y je podmnožina učební množiny, jejíž obrazy byly chybně klasifikovány s daným nastavením váhového vektoru **w**; hodnoty proměnné $\delta_x$ jsou stanoveny tak, že $\delta_x = -1$ pro $\mathbf{x} \in \omega_1$ a $\delta_x = 1$ pro $\mathbf{x} \in \omega_2$.
- ☑ součet (🚓) je zřejmě vždycky nezáporný a roven nule pokud Y je prázdná množina.
- ☑ je to funkce spojitá a po částech lineární (gradient není definován ve chvíli, kdy se mění počet chybně klasifikovaných vektorů x)

# PERCEPTRON

algoritmus výpočtu **w*** (gradientní metoda):

$$\mathbf{w}(t+1) = \mathbf{w}(t) - \rho_t \left.\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}}\right|_{\mathbf{w}=\mathbf{w}(t)}$$

**w**(t) je vektor váhových koeficientů v t-tém kroku iterace;

$\rho_t > 0$

tam kde je gradient definován je

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \sum_{\mathbf{x} \in Y} \delta_x \mathbf{x}$$

po dosazení do definičního vztahu je

$$\mathbf{w}(t+1) = \mathbf{w}(t) - \rho_t \sum_{\mathbf{x} \in Y} \delta_x \mathbf{x}$$

# PERCEPTRON

## algoritmus výpočtu **w*** - pseudokód:

- zvolte náhodně w(0)
- zvolte $\rho_0$
- t=0
- repeat

  Y={∅}

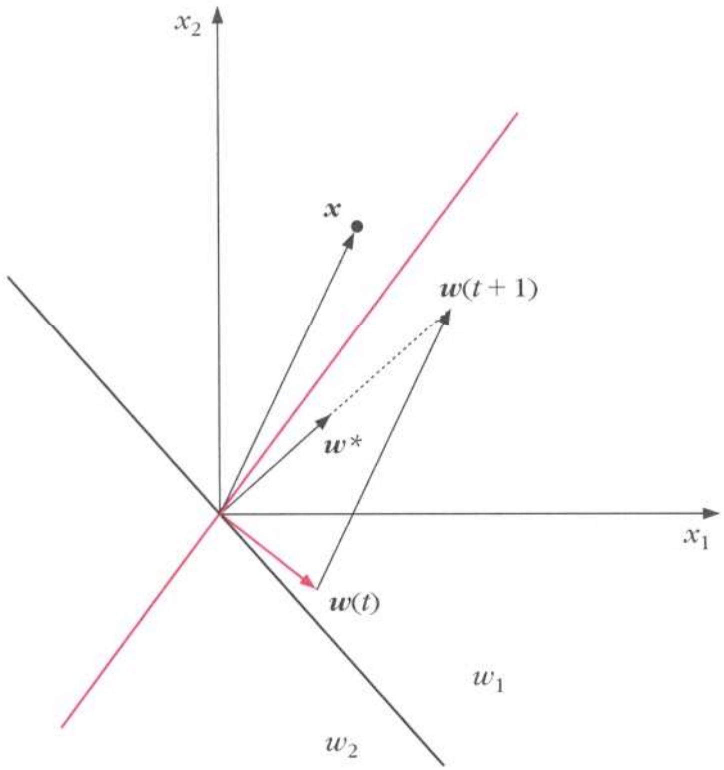
  for i=1 to N

    if $\delta_{x_i}\mathbf{w}(t)^T\mathbf{x}_i \geq 0$ then $Y = Y \cup \{\mathbf{x}_i\}$

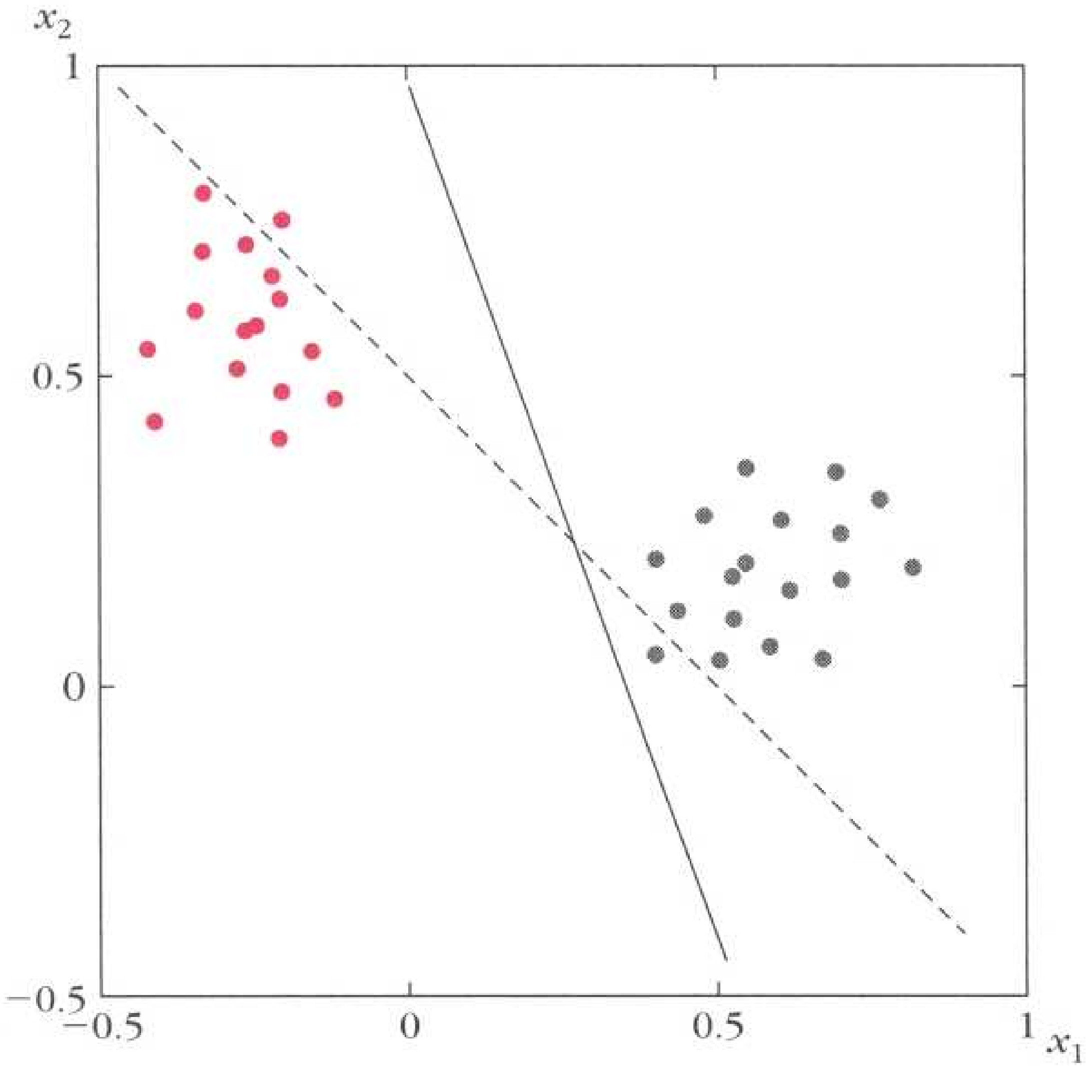
  $\mathbf{w}(t+1) = \mathbf{w}(t) - \rho_t\Sigma_{x\in Y}\delta_x\mathbf{x}$

  nastavte $\rho_t$

  t=t+1
- until Y={∅}



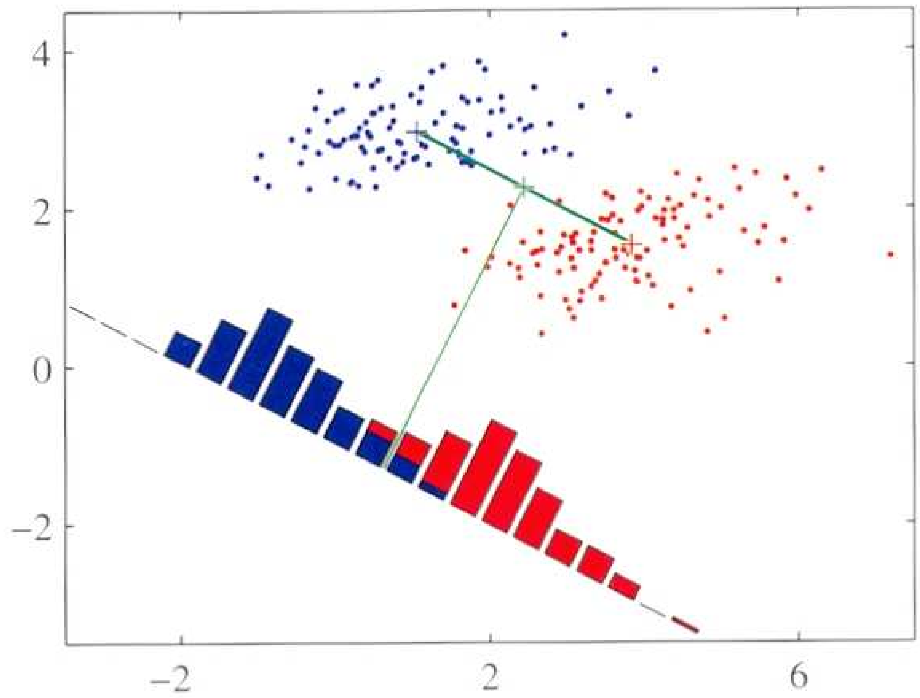
# PERCEPTRON

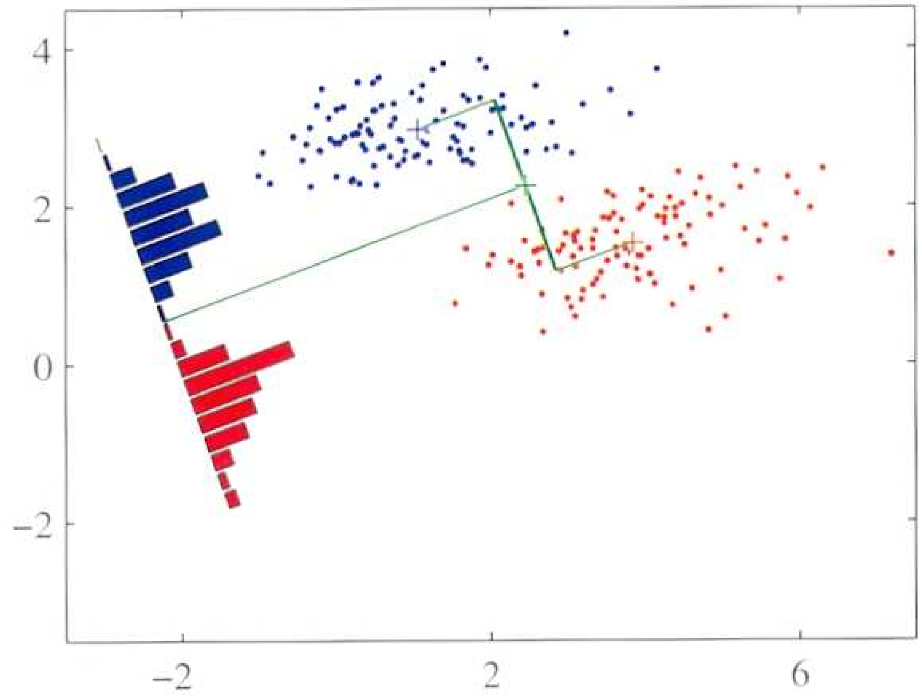


# FISHEROVA DISKRIMINACE

- ☑ redukce dimenzionality?
- ☑ nejdříve dichotomický problém:
  - ➔ předpokládejme na vstupu n-rozměrný vektor **x**, který promítneme do jednoho rozměru pomocí $y=\mathbf{w}^T\mathbf{x}$
  - ➔ projekcí do jednoho rozměru ztrácíme mnohou zajímavou informaci, ale určením prvků váhového vektoru **w** můžeme nastavit projekci, která maximalizuje separaci tříd;

# FISHEROVA DISKRIMINACE



# FISHEROVA DISKRIMINACE

- ☑ předpokládejme, že známe učební množinu $n_1$ obrazů z třídy $\omega_1$ a $n_2$ obrazů z $\omega_2$;
- ☑ střední vektory reprezentující každou třídu jsou

$$\mathbf{m}_1 = \frac{1}{n_1}\sum_{i \in \omega_1} \mathbf{x}_i \quad \text{a} \quad \mathbf{m}_2 = \frac{1}{n_2}\sum_{i \in \omega_1} \mathbf{x}_i$$

- ☑ nejjednodušší míra separace klasifikačních tříd, je separace klasifikačních průměrů, tj. stanovení w tak, aby byla maximalizována hodnota $m_2$-$m_1$=$\mathbf{w}^T(\mathbf{m}_2$-$\mathbf{m}_1)$, kde $m_r$=$\mathbf{w}^T\mathbf{m}_r$ je průměr projektovaných dat ze třídy $\omega_r$;

# FISHEROVA DISKRIMINACE

- ☑ aby hodnota $m_2$-$m_1$ neomezeně nerostla s růstem modulu w, předpokládáme jeho jednotkovou délku, tj. $\Sigma_i w_i^2=1$
- ☑ s použitím Langrangova součinitele (multiplikátoru) pro hledání vázaného extrému

$$\mathbf{w} \propto (\mathbf{m}_2 - \mathbf{m}_1)$$

**Fisherův diskriminátor**

podle Fisherova pravidla stanovíme pouze optimální směr souřadnice, na kterou promítáme obrazy klasifikovaných tříd.

abychom stanovili rozhodovací pravidlo, musíme určit hodnotu prahu $w_0$

# LANGRANGŮV SOUČINITEL

## ☑ Langragova metoda neurčitých koeficientů

Nechť f(x,y) a g(x,y) mají v okolí bodů křivky g(x,y)=0 totální diferenciál. Nechť v každém bodě křivky g(x,y)=0 je aspoň jedna z derivací $\partial g/\partial x$, $\partial g/\partial y$ různá od nuly. Má-li funkce z=f(x,y,) v bodě $[x_0,y_0]$ křivky g(x,y)=0 lokální extrém na této křivce, pak existuje taková konstanta $\lambda$, že pro funkci

$$F(x,y)=f(x,y) + \lambda.g(x,y) \qquad (\heartsuit)$$

jsou v bodě $[x_0,y_0]$ splněny rovnice

$$\partial F(x_0,y_0)/\partial x=0;\ \partial F(x_0,y_0)/\partial y=0 \qquad (🕊)$$

a samozřejmě $g(x_0,y_0)=0$ (**podmínky nutné**).

Vázané extrémy lze tedy hledat tak, že sestrojíme funkci ($\heartsuit$) a řešíme rovnice (🕊) pro neznámé $x_0,y_0$, $\lambda$ ($\lambda$ nazýváme Lagrangeův součinitel (multiplikátor)).

# LANGRANGŮV SOUČINITEL

☑ Langragova metoda neurčitých koeficientů

totální diferenciál:

Je-li $f(x,y)$ v $[x_0,y_0]$ diferencovatelná, nazývá se výraz

$$dz = (\partial f/\partial x).dx + (\partial f/\partial y).dy$$

totální diferenciál funkce $z=f(x,y)$.

# LANGRANGŮV SOUČINITEL

☑ Langragova metoda neurčitých koeficientů

**podmínky postačující:**

Sestrojme v bodě $[x_0,y_0]$ druhý diferenciál funkce (♥)

$$d^2F(x_0,y_0)= \partial^2F(x_0,y_0)/\partial x^2+2\partial^2F(x_0,y_0)/\partial x\ \partial x\ +\partial^2F(x_0,y_0)/\partial y^2$$ (✉)

Jestliže pro všechny body $[x_0+dx,y_0+dy]$ z určitého okolí bodu $[x_0,y_0]$ takové, že $g(x_0+dx,y_0+dy)=0$ a že dx a dy nejsou zároveň rovny nule, je (✉) kladné, resp. záporné, pak je v bodě $[x_0,y_0]$ vázaný lokální extrém, a to minimum (resp. maximum).

# LANGRANGŮV SOUČINITEL

## ☑ Langragova metoda neurčitých koeficientů

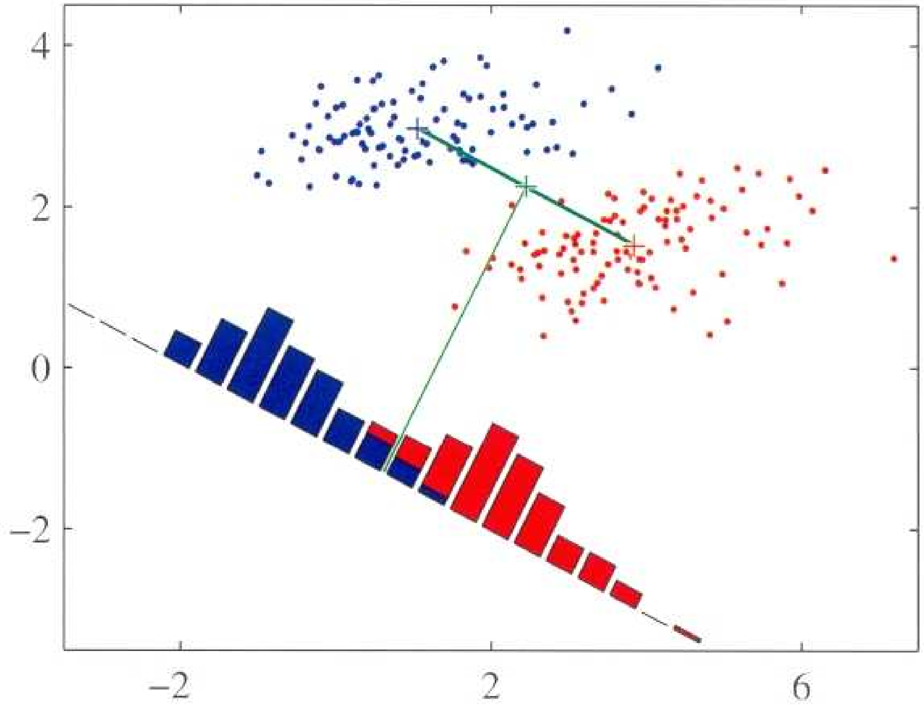
Obdobně se řeší úloha najít vázané extrémy funkce několika proměnných, např. nutná podmínka k existenci lokálního extrému funkce $w=f(x,y,z,u,v)$ při podmínkách $F_1(x,y,z,u,v)$, $F_2(x,y,z,u,v)$ je splnění rovnic

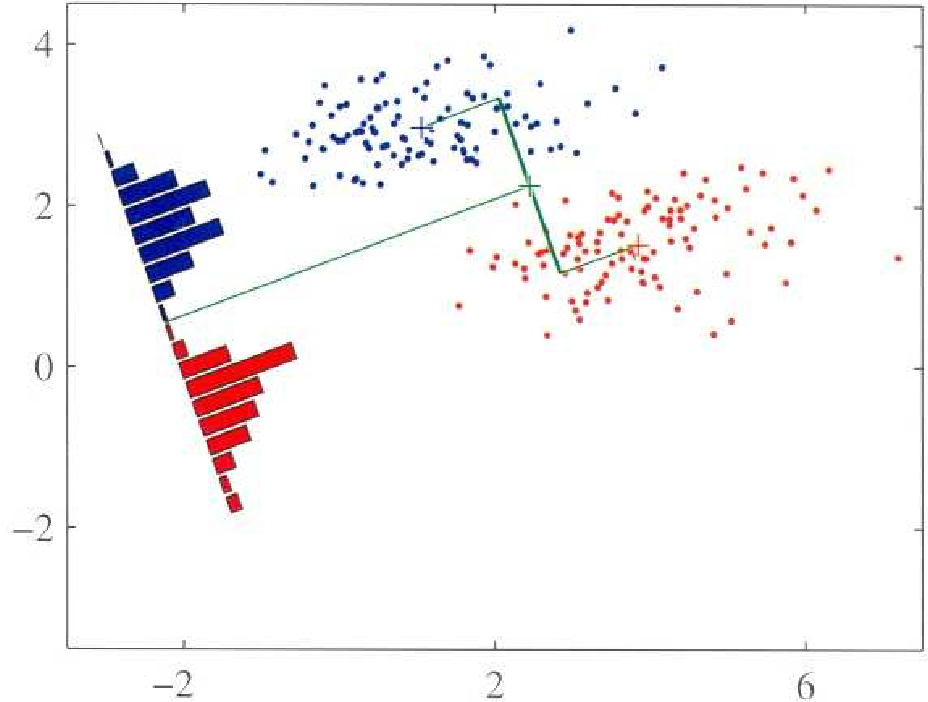
$\partial G/\partial x=0$, $\partial G/\partial y=0$, $\partial G/\partial z=0$, $\partial G/\partial u=0$, $\partial G/\partial v=0$, $F_1=0$ a $F_2=0$,

kde $G= f+ \lambda_1 F_1+\lambda_2 F_2$, tj. soustava 7 rovnic pro 7 neznámých.

# FISHEROVA DISKRIMINACE

problém:




řešení:
nejen maximální vzdálenost tříd, ale současně i minimální rozptyl uvnitř tříd

# FISHEROVA DISKRIMINACE

- ☑ rozptyl transformovaných dat ze třídy $\omega_1$ je dána $$s_r^2 = \sum_{i \in \omega_1} (y_i - m_i)^2$$
  kde $y_i = \mathbf{w}^T \mathbf{x}_i$;
- ☑ celkový rozptyl uvnitř klasifikačních tříd z celé báze dat jednoduše součtem $s_1^2 + s_2^2$

# FISHEROVA DISKRIMINACE

- Fisherovo kritérium:

$$J(\mathbf{w}) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2}$$

- po dosazení maticově:

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

kde $\mathbf{S}_B$ je kovarianční matice mezi třídami

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

a $\mathbf{S}_W$ je matice celkové kovariance uvnitř tříd

$$\mathbf{S}_W = \sum_{i \in \omega_1} (\mathbf{x}_i - \mathbf{m}_1)(\mathbf{x}_i - \mathbf{m}_1)^T + \sum_{i \in \omega_2} (\mathbf{x}_i - \mathbf{m}_2)(\mathbf{x}_i - \mathbf{m}_2)^T$$

# FISHEROVA DISKRIMINACE

☑ maximální J(**w**) určíme po derivaci (🔔) podle **w** tehdy, když platí

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

z toho pak

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

**Fisherův diskriminátor**

směr vektoru $\mathbf{m}_2-\mathbf{m}_1$ je na rozdíl od původního případu modifikován maticí $\mathbf{S}_W$;

pokud je kovariance uvnitř tříd izotropní (rozptyl je týž ve všech směrech), $\mathbf{S}_W$ je úměrná jednotkové matici a w má opět směr vektoru $\mathbf{m}_2-\mathbf{m}_1$

# ALGORITMUS PODPŮRNÝCH VEKTORŮ

**Algoritmus podpůrných vektorů (Support Vector Machine** - SVM) je metoda strojového učení. SVM hledá nadrovinu, která v prostoru příznaků optimálně rozděluje trénovací data.

Optimální nadrovina je taková, že body leží v opačných poloprostorech a hodnota minima vzdáleností bodů od roviny je co největší. Jinými slovy, okolo nadroviny je na obě strany co nejširší pruh bez bodů.

Na popis nadroviny stačí pouze nejbližší body, kterých je obvykle málo - tyto body se nazývají *podpůrné vektory* (angl. support vectors) a odtud název metody. Tato metoda je ze své přirozenosti binární, tedy rozděluje data do dvou množin. Rozdělující nadrovina je lineární funkcí v prostoru příznaků.

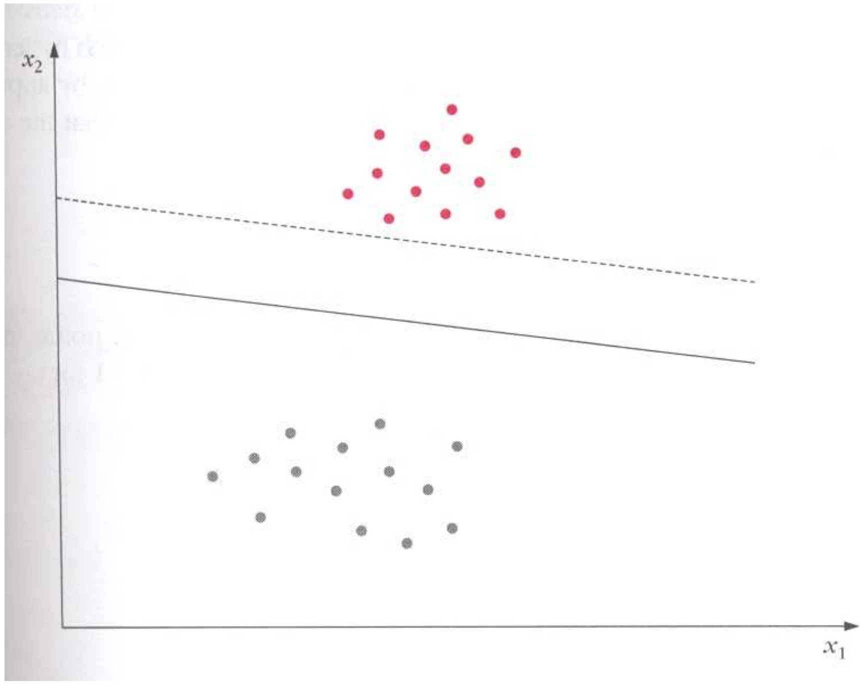
# SEPARABILNÍ TŘÍDY

mějme v učební množině obrazy $\mathbf{x}_i$, i=1,2,…,n, ze dvou lineárně separabilních klasifikačních tříd $\omega_1$ a $\omega_2$

cílem je určení parametrů definující hranici

$$y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = 0,$$

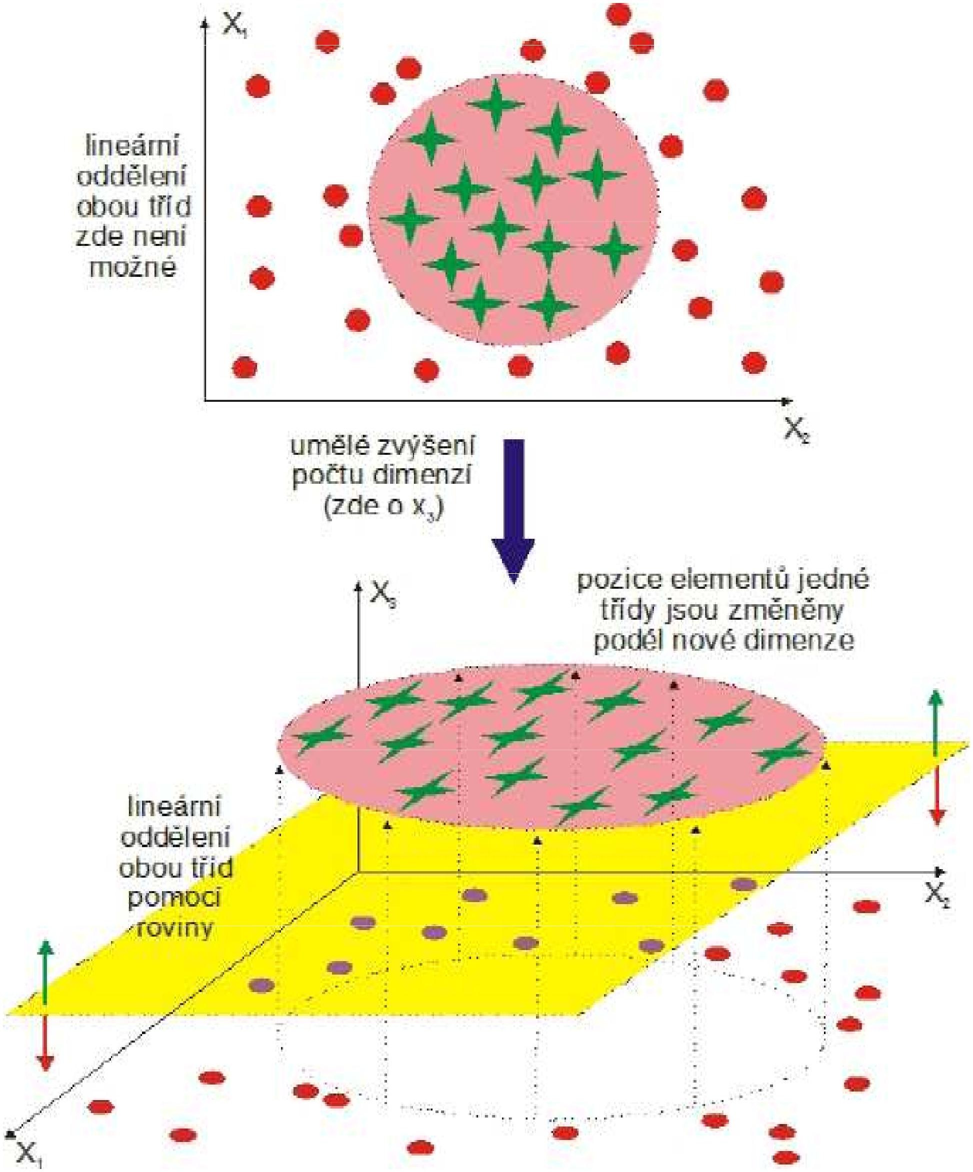
jejíž pomocí klasifikátor správně zařadí všechny obrazy z učební množiny



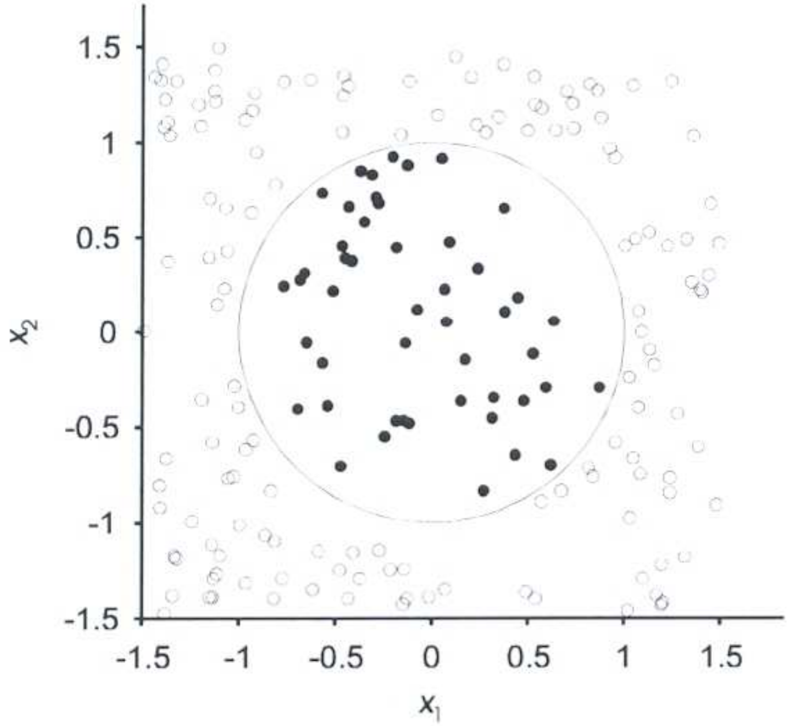
# ALGORITMUS PODPŮRNÝCH VEKTORŮ

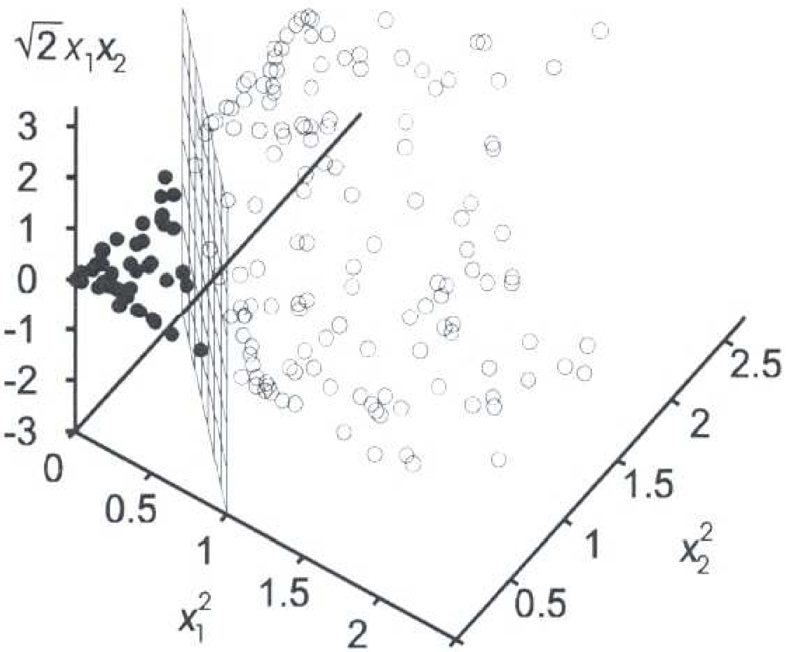
- Důležitou součástí techniky **SUPPORT VECTOR MACHINES** je jádrová transformace (angl. *kernel transformation*) prostoru příznaků dat do prostoru transformovaných příznaků typicky vyšší dimenze. Tato jádrová transformace umožňuje převést původně lineárně neseparovatelnou úlohu na úlohu lineárně separovatelnou, na kterou lze dále aplikovat optimalizační algoritmus pro nalezení rozdělující nadroviny.
- Používají se různé jádrové transformace. Intuitivně, vyjadřují podobnost dat, tj. svých dvou vstupních argumentů.
- Výhodou této metody (a jiných metod založených na jádrové transformaci) je, že transformace se dá definovat pro různé typy objektů, nejen body v $R^n$. Např. pro grafy, stromy, posloupnosti DNA ...

# ALGORITMUS PODPŮRNÝCH VEKTORŮ



# ALGORITMUS PODPŮRNÝCH VEKTORŮ




dvourozměrný prostor s oddělovací hranicí ve tvaru $x_1^2 + x_2^2 \leq 1$

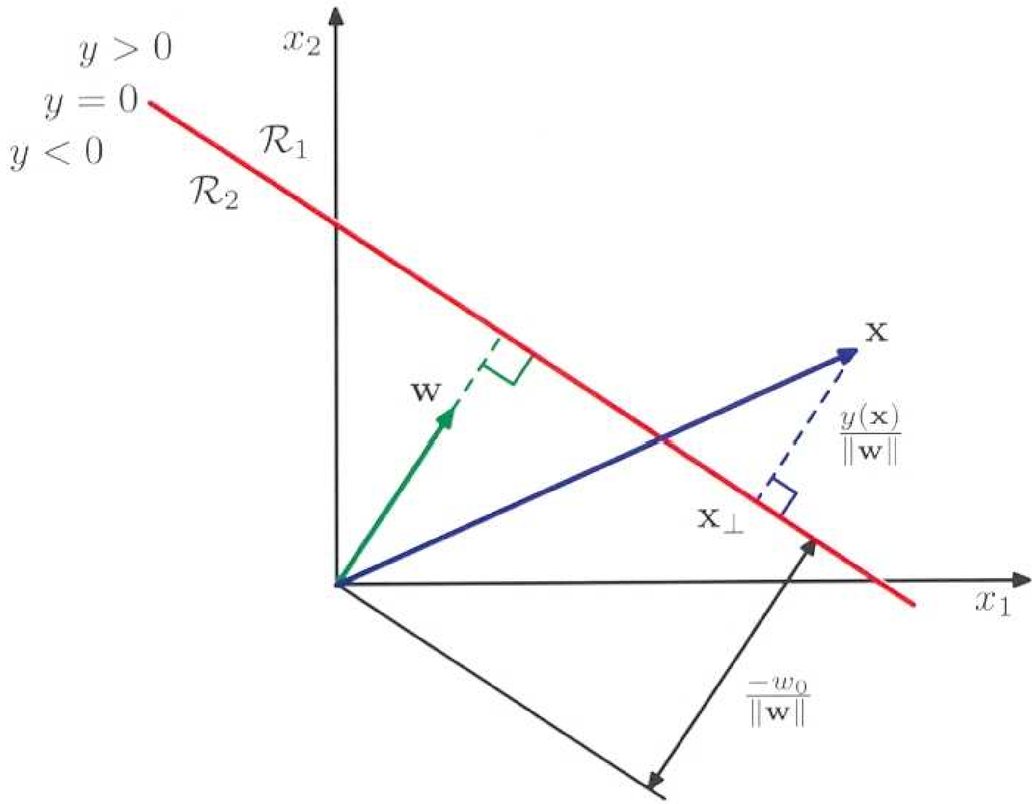
tatáž situace zobrazená do trojrozměrného prostoru $(x_1^2, x_2^2, \sqrt{2}x_1x_2)$ – **kruhová hranice se stane lineární**

# ALGORITMUS PODPŮRNÝCH VEKTORŮ
## SEPARABILNÍ TŘÍDY

☑ připomenutí:

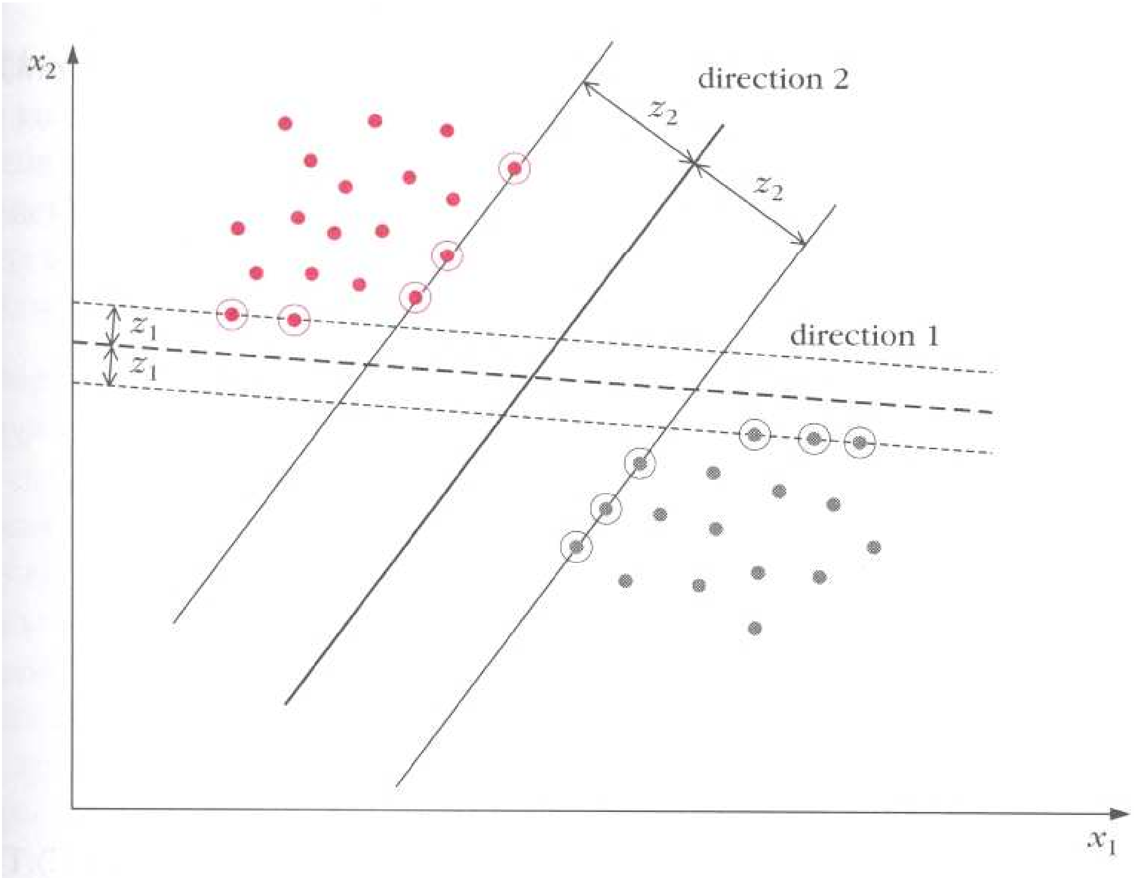
vzdálenost jakéhokoliv bodu od klasifikační hranice je

$$d = \frac{|y(\mathbf{x})|}{\|\mathbf{w}\|}$$



# ALGORITMUS PODPŮRNÝCH VEKTORŮ
## SEPARABILNÍ TŘÍDY

- ☑ určeme hodnoty váhového vektoru **w** a $w_0$ tak, aby hodnota y(**x**) v nejbližším bodě třídy $\omega_1$ byla rovna 1 a pro $\omega_2$ rovna -1



# ALGORITMUS PODPŮRNÝCH VEKTORŮ
## SEPARABILNÍ TŘÍDY

☑ máme „ochranné“ klasifikační pásmo o šířce

$$\frac{1}{\|\mathbf{w}\|}+\frac{1}{\|\mathbf{w}\|}=\frac{2}{\|\mathbf{w}\|}$$

a chceme
$$\mathbf{w}^T\mathbf{x}+w_0 \geq 1 \quad \text{pro } \forall \mathbf{x} \in \omega_1$$
$$\mathbf{w}^T\mathbf{x}+w_0 \geq -1 \quad \text{pro } \forall \mathbf{x} \in \omega_2$$

nebo také - chceme najít minimální

$$J(\mathbf{w}, w_0)=\frac{1}{2}\|\mathbf{w}\|^2$$

za předpokladu, že

$$t_i(\mathbf{w}^T\mathbf{x}+w_0) \geq 1, \quad i=1,2,\ldots,n$$

kde $t_i$ =+1 pro $\omega_1$ a $t_i$ =-1 pro $\omega_2$

(minimalizace normy maximalizuje klasifikační pásmo)

# ALGORITMUS PODPŮRNÝCH VEKTORŮ
## SEPARABILNÍ TŘÍDY

- nelineární kvadratická optimalizační úloha se soustavou podmínek formulovaných pomocí lineárních nerovností
- Karushovy-Kuhnovy-Tuckerovy podmínky praví, že pro to musí být splněno

$$\frac{\partial}{\partial \mathbf{w}} L(\mathbf{w}, w_0, \boldsymbol{\lambda}) = \mathbf{0}$$

$$\frac{\partial}{\partial w_0} L(\mathbf{w}, w_0, \boldsymbol{\lambda}) = 0$$

$$\lambda_i \geq 0, i = 1, 2, \ldots, n$$

$$\lambda_i [t_i (\mathbf{w}^T \mathbf{x} + w_0) - 1] = 0, i = 1, 2, \ldots, n,$$

kde $\boldsymbol{\lambda}$ je vektor Langrangových součinitelů a L(**w**,w0, $\boldsymbol{\lambda}$) je Lagrangova funkce definována vztahem

$$L(\mathbf{w}, w_0, \boldsymbol{\lambda}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^{n} \lambda_i [t_i (\mathbf{w}^T \mathbf{x}_i + w_0) - 1]$$

# ALGORITMUS PODPŮRNÝCH VEKTORŮ
## SEPARABILNÍ TŘÍDY

- ☑ když se všechny vztahy z předcházející strany dají dohromady dostaneme

$$\mathbf{w} = \sum_{i=1}^{n} \lambda_i t_i \mathbf{x}_i$$

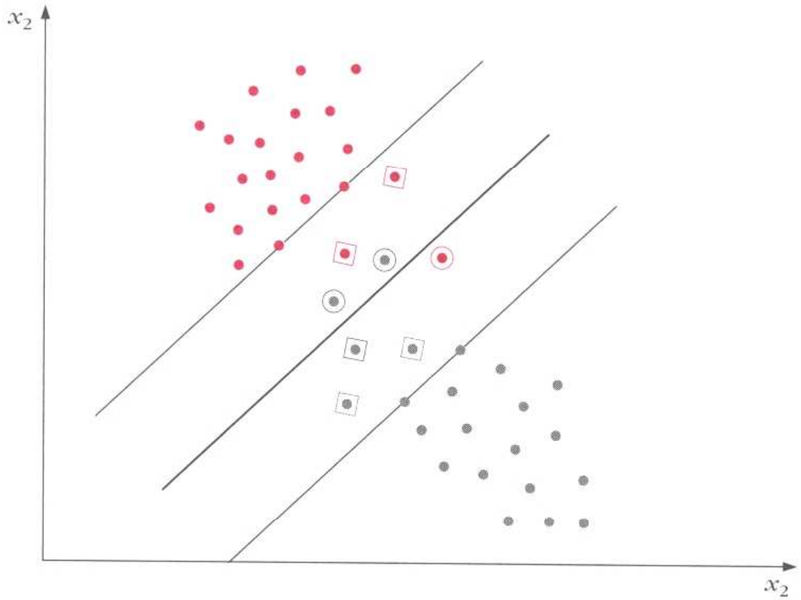
**podpůrné vektory**

$$\sum_{i=1}^{n} \lambda_i t_i = 0$$

# ALGORITMUS PODPŮRNÝCH VEKTORŮ
## NESEPARABILNÍ TŘÍDY

- ☑ stále ale platí, že klasifikační „ochranné" pásmo je definováno dvěma paralelními „nadrovinami" definovanými

$$\mathbf{w}^T\mathbf{x} + w_0 = \pm 1$$




- ☑ obrazy z trénovací množiny patří do následujících tří kategorií:
  - ➔ obraz leží **vně** pásma a je **správně** klasifikován [platí podmínka $t_i(\mathbf{w}^T\mathbf{x}+w_0)\geq 1$ i=1,2,…,n];
  - ➔ obraz leží **uvnitř** pásma a je **správně** klasifikován (čtverečky) [platí pro ně $0\leq t_i(\mathbf{w}^T\mathbf{x}+w_0)<1$];
  - ➔ obraz je chybně klasifikován (kolečka) [platí pro něj $t_i(\mathbf{w}^T\mathbf{x}+w_0)<0$]

# ALGORITMUS PODPŮRNÝCH VEKTORŮ
## NESEPARABILNÍ TŘÍDY

- ☑ všechny tři kategorie obrazů mohou být řešeny na základě pro daný typ specifických podmínek

$$t_i(\mathbf{w}^T\mathbf{x}+w_0)\geq 1-\xi_i$$

pomocí nově zavedených proměnných $\xi_i$ (tzv. volné proměnné - *slack variables*).

První kategorie je pro $\xi_i=0$, druhá $0<\xi_i\leq 1$ a třetí pro $\xi_i>1$.

Cílem návrhu v tomto případě je vytvořit co nejširší „ochranné" pásmo, ale současně minimalizovat počet obrazů s $\xi_i>0$, což vyjadřuje kritérium se ztrátovou funkcí

$$J(\mathbf{w},w_0,\boldsymbol{\xi})=\frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{n}I(\xi_i)$$

kde $\boldsymbol{\xi}$ je vektor parametrů $\xi_i$ a

$$I(\xi_i)=\begin{cases}1 & \xi_i>0\\ 0 & \xi_i=0\end{cases}$$

C je kladná korekční konstanta, která váhuje vliv obou členů v uvedeném vztahu.

# ALGORITMUS PODPŮRNÝCH VEKTORŮ
## NESEPARABILNÍ TŘÍDY

- ☑ optimalizace je obtížná, protože ztrátová funkce je nespojitá (díky funkci I(•)). V takových případech se proto používá náhradní ztrátová funkce

$$J(\mathbf{w}, w_0, \boldsymbol{\xi}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$

- ☑ a cílem návrhu je minimalizovat $J(\mathbf{w},w_0,\boldsymbol{\xi})$ za podmínek, že

$$t_i(\mathbf{w}^T\mathbf{x}+w_0)\geq 1-\xi_i \text{ a } \xi_i \geq 0,\ i=1,2,\ldots,n.$$

- ☑ Problém lze opět řešit pomocí Langrangeovy funkce

$$L(\mathbf{w}, w_0, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\mu_i\xi_i - \sum_{i=1}^{n}\lambda_i[t_i(\mathbf{w}^T\mathbf{x}_i + w_0) - 1 + \xi_i]$$

# ALGORITMUS PODPŮRNÝCH VEKTORŮ
## NESEPARABILNÍ TŘÍDY

- ☑ příslušné Karushovy-Kuhnovy-Tuckerovy podmínky jsou

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{0} \text{ nebo } \mathbf{w} = \sum_{i=1}^{n} \lambda_i t_i \mathbf{x}_i;$$

$$\frac{\partial L}{\partial w_0} = 0 \text{ nebo } \sum_{i=1}^{n} \lambda_i t_i \mathbf{x}_i = 0;$$

$$\frac{\partial L}{\partial \xi_i} = 0 \text{ nebo } C - \mu_i - \lambda_i = 0, i = 1, 2, \ldots, n;$$

$$\lambda_i \left[ t_i (\mathbf{w}^T \mathbf{x} + w_0) - 1 + \xi_i \right] = 0, i = 1, 2, \ldots, n;$$

$$\mu_i \xi_i = 0, i = 1, 2, \ldots, n;$$

$$\mu_i \geq 0, \lambda_i \geq 0, i = 1, 2, \ldots, n;$$

# ALGORITMUS PODPŮRNÝCH VEKTORŮ
## NESEPARABILNÍ TŘÍDY

- ☑ z čehož platí požadavek na maximalizaci $L(\mathbf{w}, w_0, \boldsymbol{\lambda}, \boldsymbol{\xi}, \boldsymbol{\mu})$ za podmínek

$$\mathbf{w} = \sum_{i=1}^{n} \lambda_i t_i \mathbf{x}_i;$$

$$\sum_{i=1}^{n} \lambda_i t_i = 0;$$

$$C - \mu_i - \lambda_i = 0,$$

$$\mu_i \geq 0, \lambda_i \geq 0, i = 1, 2, \ldots, n;$$

## KLASIFIKACE PODLE MINIMÁLNÍ VZDÁLENOSTI

- reprezentativní obrazy klasifikačních tříd - etalony
- je-li v obrazovém prostoru zadáno R poloh etalonů vektory $\mathbf{x}_{1E}, \mathbf{x}_{2E}, \ldots, \mathbf{x}_{RE}$, zařadí klasifikátor podle minimální vzdálenosti klasifikovaný obraz **x** do té třídy, jejíž etalon má od bodu **x** minimální vzdálenost. Rozhodovací pravidlo je určeno vztahem

$$d(\mathbf{x}) = \left\| \mathbf{x}_{rE} - \mathbf{x} \right\| = \min_{\forall s} \left\| \mathbf{x}_{sE} - \mathbf{x} \right\|$$

# PRINCIPY KLASIFIKACE

- ☑ pomocí diskriminačních funkcí – funkcí, které určují míru příslušnosti k dané klasifikační třídě;
- ☑ pomocí definice hranic mezi jednotlivými třídami a logických pravidel;
- ☑ pomocí vzdálenosti od reprezentativních obrazů (etalonů) klasifikačních tříd;
- ☑ pomocí ztotožnění s etalony;

# SOUVISLOSTI MEZI PRINCIPY KLASIFIKACE

- ☑ uvažme případ dvou tříd reprezentovaných etalony $\mathbf{x}_{1E} = (x_{11E}, x_{12E})$ a $\mathbf{x}_{2E} = (x_{21E}, x_{22E})$ ve dvoupříznakovém euklidovském prostoru;
- ☑ vzdálenost mezi obrazem $\mathbf{x} = (x_1, x_2)$ a libovolným z obou etalonů je pak definována

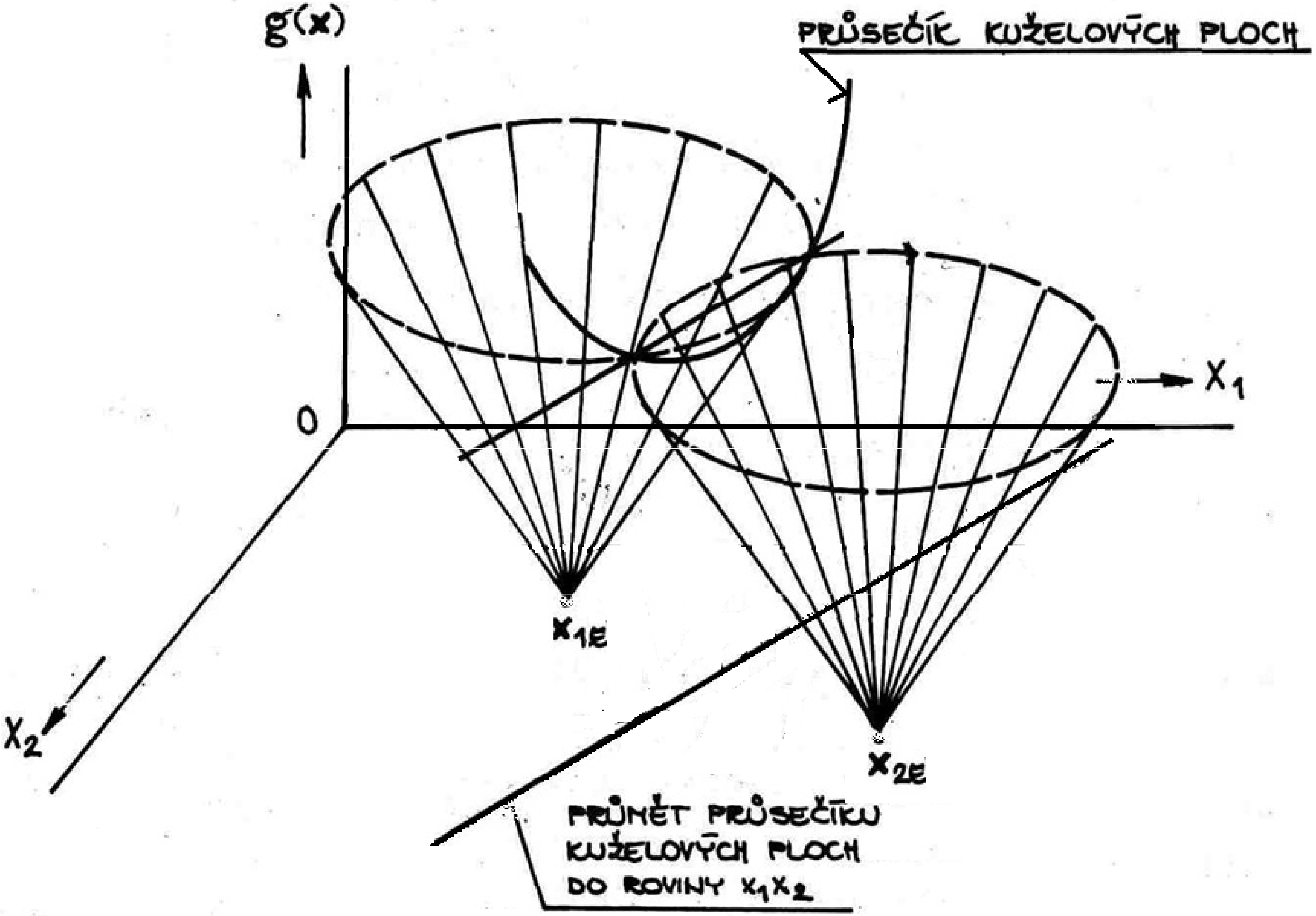
$$v(\mathbf{x}_{sE}, \mathbf{x}) = \left\| \mathbf{x}_{sE} - \mathbf{x} \right\| = \sqrt{(x_{s1E} - x_1)^2 + (x_{s2E} - x_2)^2}$$

- ☑ hledáme menší z obou vzdáleností, tj. $\min_{s=1,2} v(\mathbf{x}_{sE}, \mathbf{x})$, ale také $\min_{s=1,2} v^2(\mathbf{x}_{sE}, \mathbf{x})$;

$$\min_{\forall s} v(\mathbf{x}_{sE}, \mathbf{x}) \approx \min_{\forall s} v^2(\mathbf{x}_{sE}, \mathbf{x}) = \min_{\forall s} \left( (x_{s1E} - x_1)^2 + (x_{s1E} - x_1)^2 \right) =$$

$$\min_{\forall s} \left( x_1^2 + x_2^2 - 2[x_{s1E} x_1 + x_{s2E} x_2 - (x_{s1E}^2 + x_{s2E}^2)/2] \right)$$

# SOUVISLOSTI MEZI PRINCIPY KLASIFIKACE



# SOUVISLOSTI MEZI PRINCIPY KLASIFIKACE

- ☑ diskriminační kuželové plochy se protínají v parabole a její průmět do obrazové roviny je přímka definovaná vztahem

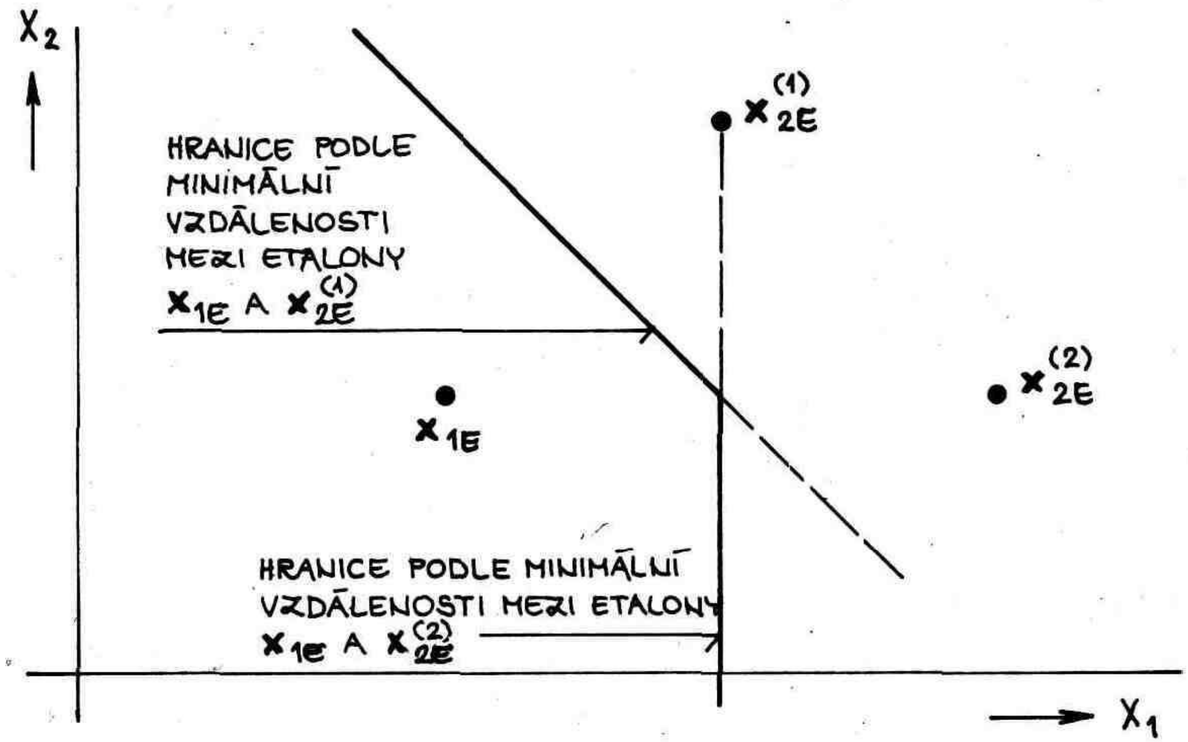
$$x_1(x_{11E} - x_{21E}) + x_2(x_{12E} - x_{22E}) - (x^2_{12E} + x^2_{11E} - x^2_{21E} - x^2_{22E})/2 = 0$$

Tato hraniční přímka mezi klasifikačními třídami je vždy kolmá na spojnici obou etalonů a tuto spojnici půlí

$$\Downarrow$$

klasifikátor pracující na základě kritéria minimální vzdálenosti je „ekvivalentní“ lineárnímu klasifikátoru s diskriminačními funkcemi.

# SOUVISLOSTI MEZI PRINCIPY KLASIFIKACE




- ☑ Klasifikace podle minimální vzdálenosti s třídami reprezentovanými více etalony je „ekvivalentní“ klasifikaci podle diskriminační funkce s po částech lineární hraniční plochou

# SOUVISLOSTI MEZI PRINCIPY KLASIFIKACE

## SHRNUTÍ

- ☑ Hranice mezi klasifikačními třídami jsou dány průmětem diskriminačních funkcí do obrazového prostoru.
- ☑ Klasifikace podle minimální vzdálenosti definuje hranici, která je kolmá na spojnici etalonů klasifikačních tříd a půlí ji.
- ☑ Princip klasifikace dle minimální vzdálenosti vede buď přímo, nebo prostřednictvím využití metrik podobnosti k definici diskriminačních funkcí a ty dle prvního ze zde uvedených pravidel k určení hranic mezi klasifikačními třídami.

Příprava nových učebních materiálů oboru Matematická biologie

je podporována projektem ESF

č. CZ.1.07/2.2.00/28.0043

**„Interdisciplinární rozvoj studijního oboru Matematická biologie"**

EVROPSKÁ UNIE

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ