# ANALÝZA A KLASIFIKACE DAT

prof. Ing. Jiří Holčík, CSc.

EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# IV. LINEÁRNÍ KLASIFIKACE

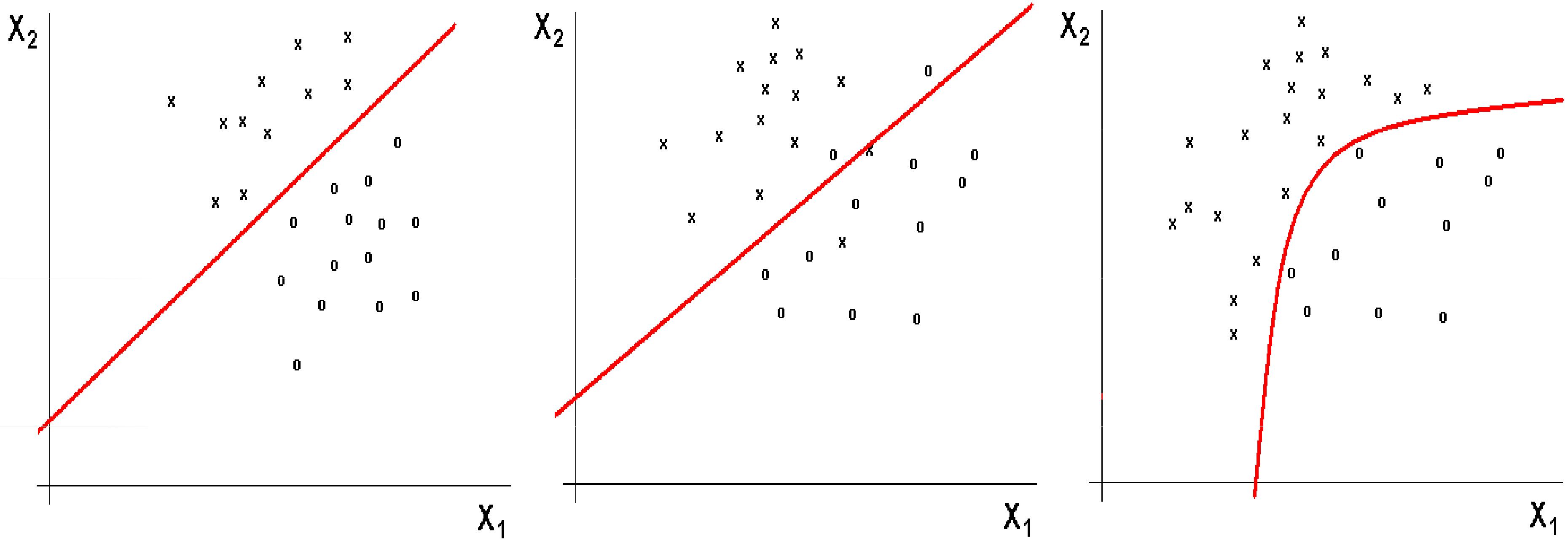
MU

IBA

# 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

# DICHOTOMICKÁ ÚLOHA
## PRINCIP

nejjednodušší realizace hraniční plochy je lineární funkcí

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

$\mathbf{w}$ 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

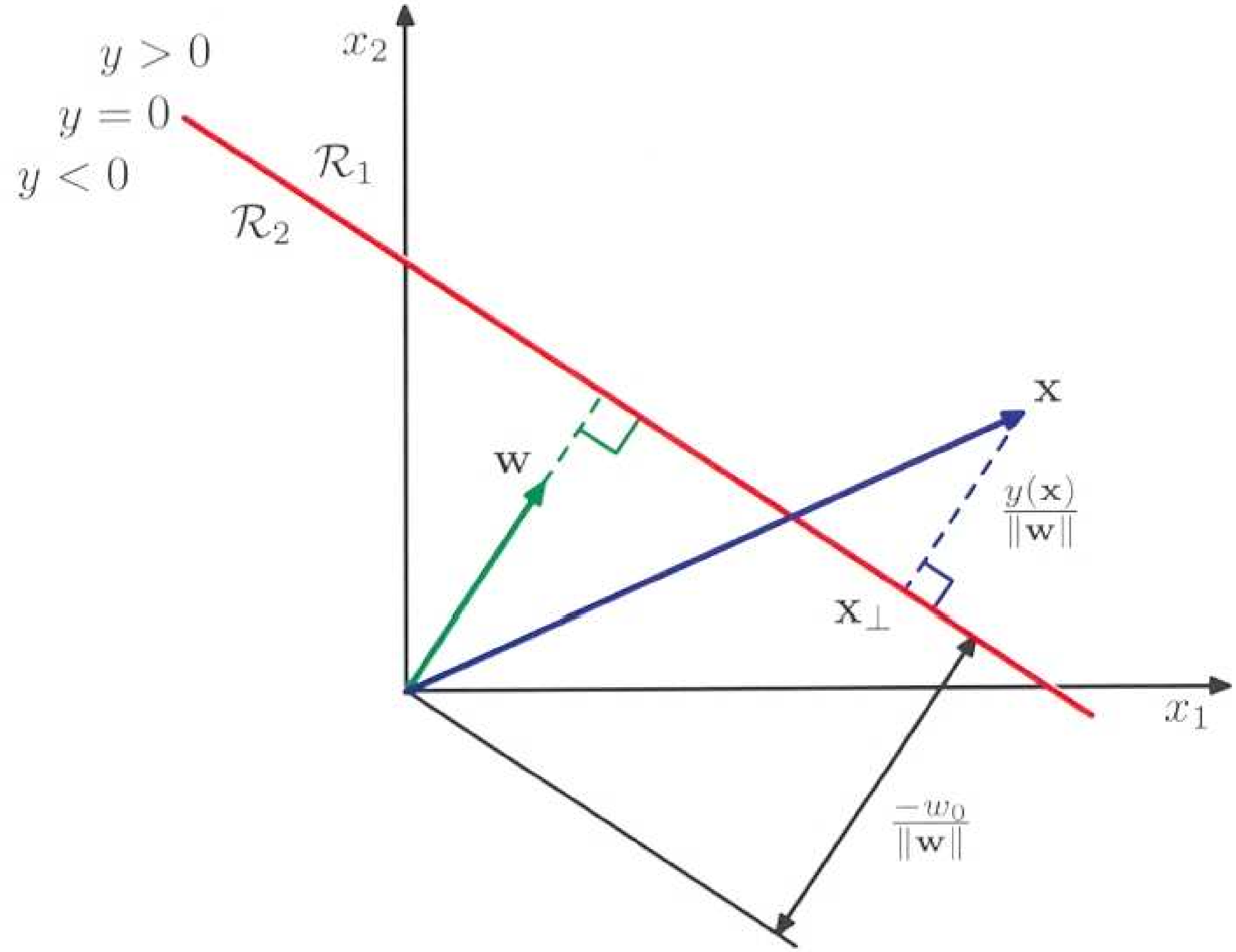
$$y(\mathbf{x}) = \tilde{\mathbf{w}}^T . \tilde{\mathbf{x}}$$

# DICHOTOMICKÁ ÚLOHA
## ZÁKLADNÍ VLASTNOSTI

- ☑ 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

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

# DICHOTOMICKÁ ÚLOHA
## ZÁKLADNÍ VLASTNOSTI



# DICHOTOMICKÁ ÚLOHA
## ZÁKLADNÍ VLASTNOSTI

$y(\mathbf{x})$ 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

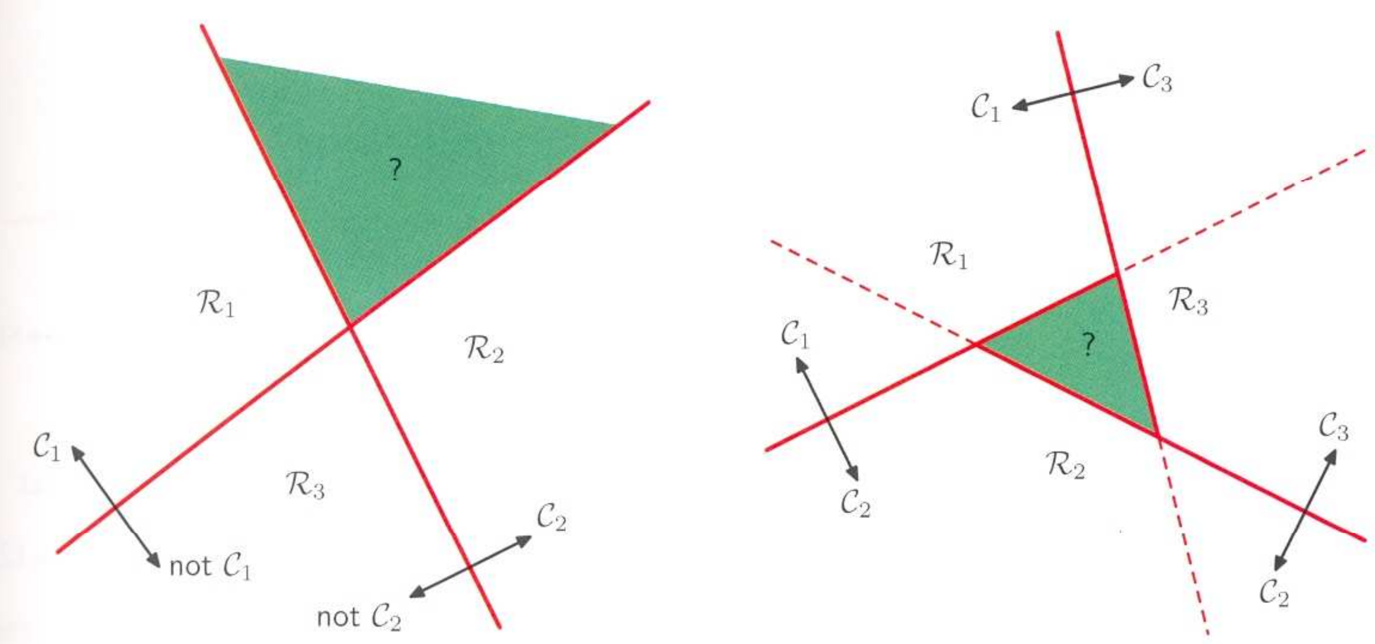
$$d = \frac{y(\mathbf{x})}{\|\mathbf{w}\|}$$

# 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

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

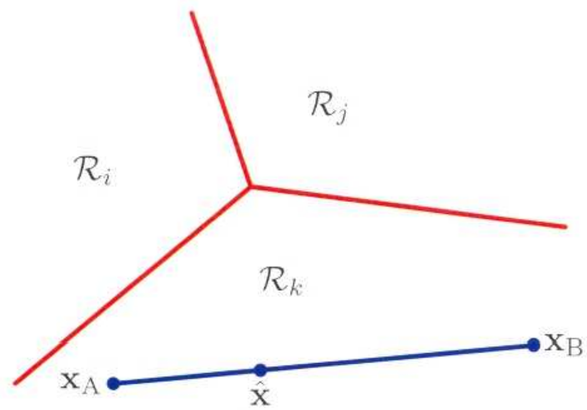
do r-té třídy $\omega_r$ zařadíme obraz $\mathbf{x}$ za předpokladu, že

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

klasifikační hranice je průmět průsečíku

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

takto definovaný klasifikační prostor je vždy spojitý a konvexní



# METODY STANOVENÍ KLASIFIKAČNÍCH HRANIC

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

# 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

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

kde $\tilde{\mathbf{W}}^{\mathrm{T}}$ je matice, jejíž r-tý sloupec zahrnuje n+1 dimenzionální vektor $\tilde{\mathbf{w}}_r = (w_0, \mathbf{w}_r^{\mathrm{T}})$ a $\tilde{\mathbf{x}} = (x_0, \mathbf{x}^{\mathrm{T}})$

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

# METODA NEJMENŠÍCH ČTVERCŮ

pokud máme učební množinu vyjádřenou $\{\mathbf{x}_i, \mathbf{t}_i\}$, i=1,…,n a i-tý řádek matice **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

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

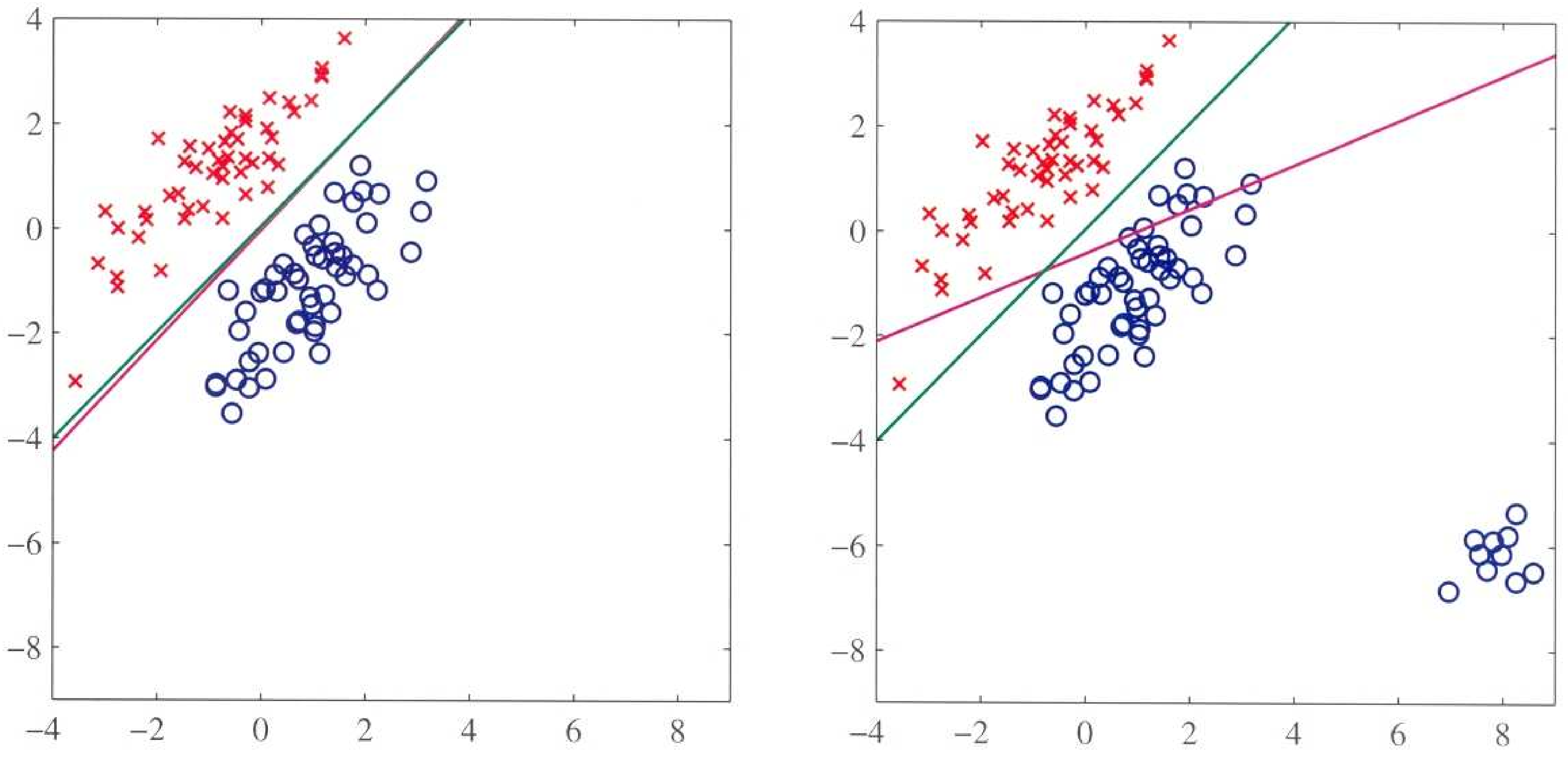
Derivací podle $\tilde{\mathbf{W}}$, kterou položíme rovno nule dostáváme

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

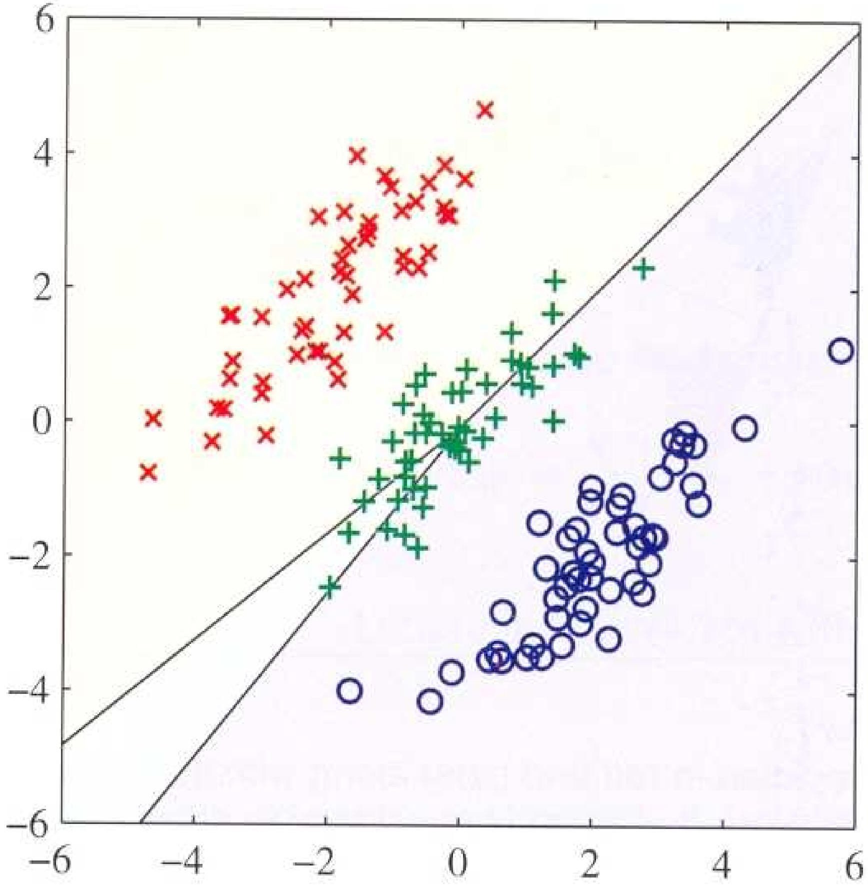
kde $\tilde{\mathbf{X}}^{\Im}$ je tzv. pseudoinverzní matice k matici $\tilde{\mathbf{X}}$ . Diskriminační funkce pak jsou ve tvaru

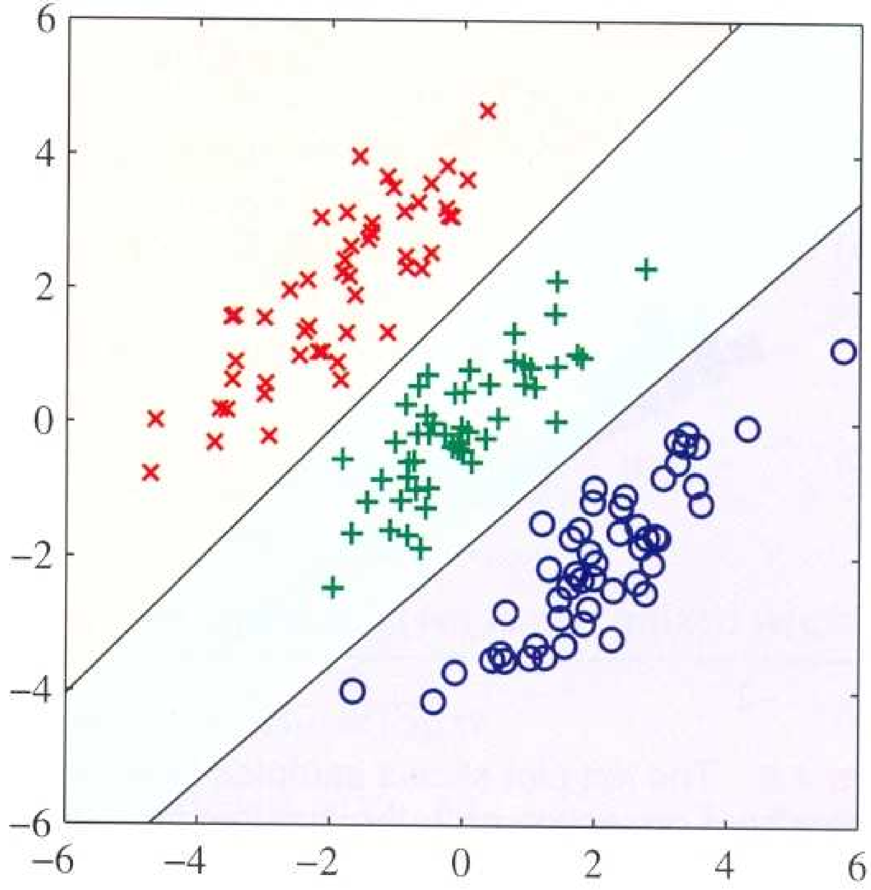
$$\mathbf{g}(\mathbf{x}) = \tilde{\mathbf{W}}^T.\tilde{\mathbf{x}} = \mathbf{T}^T(\tilde{\mathbf{X}}^{\Im})^T.\tilde{\mathbf{x}}$$

# METODA NEJMENŠÍCH ČTVERCŮ



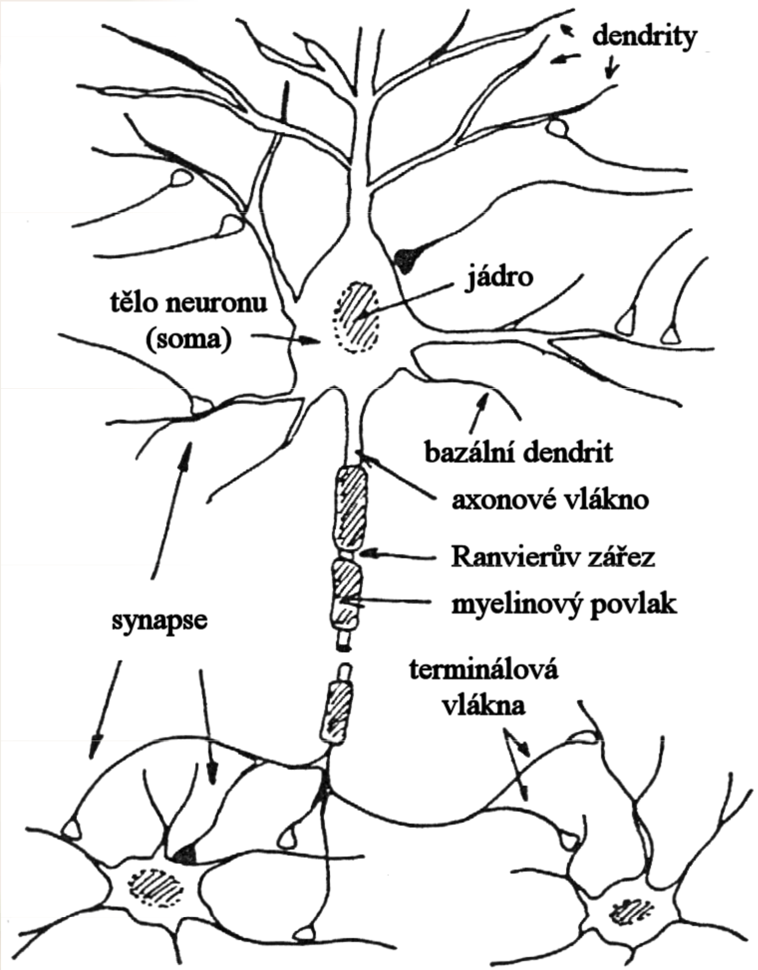
# METODA NEJMENŠÍCH ČTVERCŮ

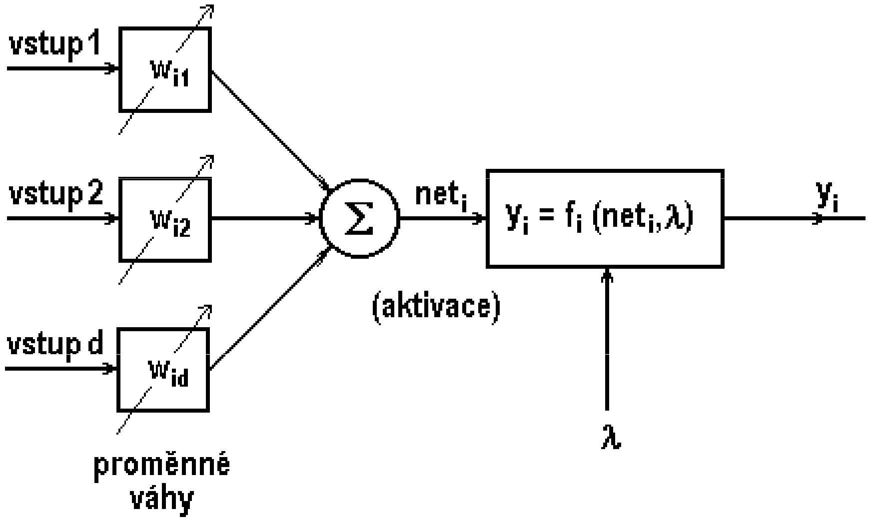


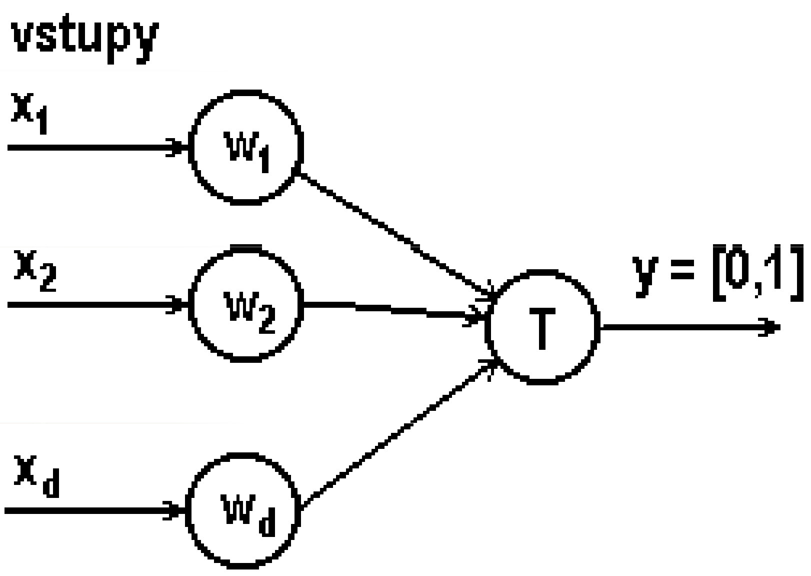
# PERCEPTRON

## 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 |

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

*Lineární model neuronu s prahem*

# PERCEPTRON

- předpokládejme, že

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

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

- 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={∅}

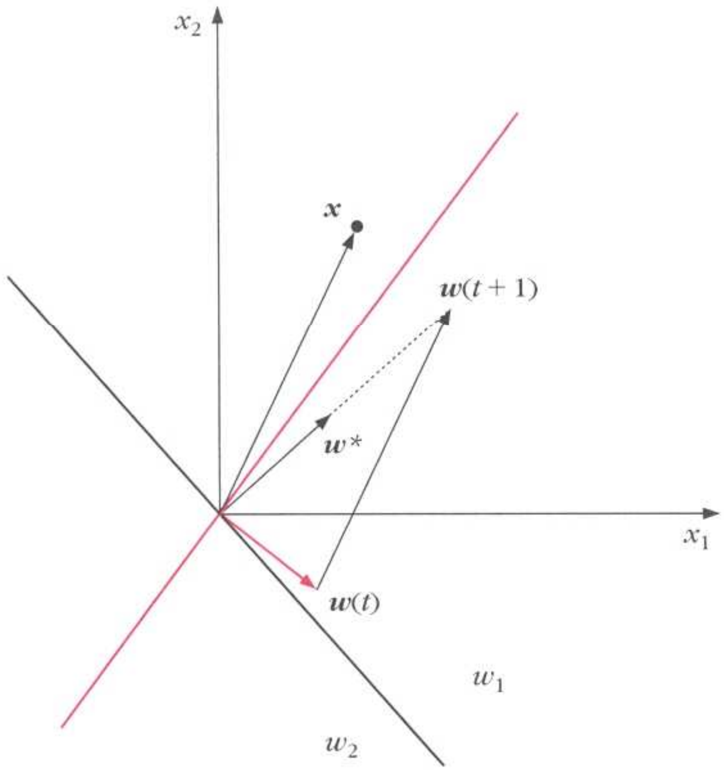
  for i=1 to N

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

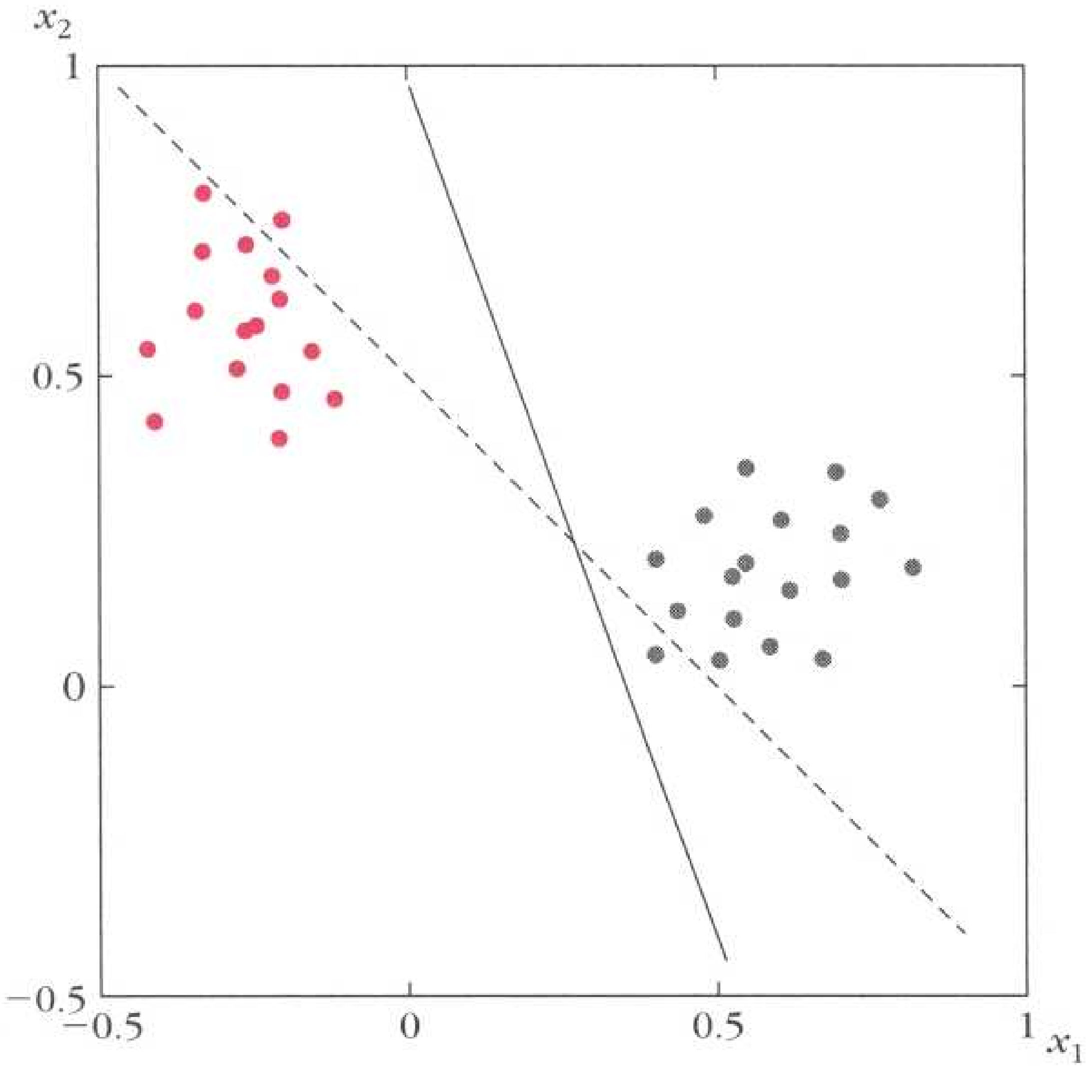
  $\mathbf{w}(t+1) = \mathbf{w}(t) - \rho_t \Sigma_{x \in Y} \delta_x \mathbf{x}$

  nastavte $\rho_t$

  t=t+1
- until Y={∅}



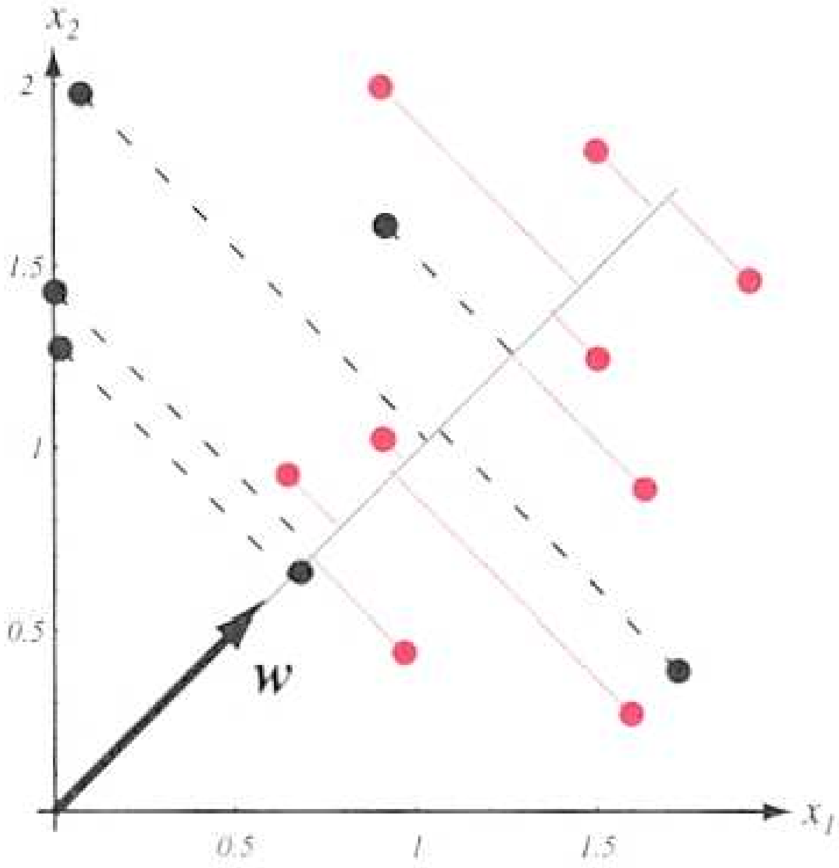
# PERCEPTRON

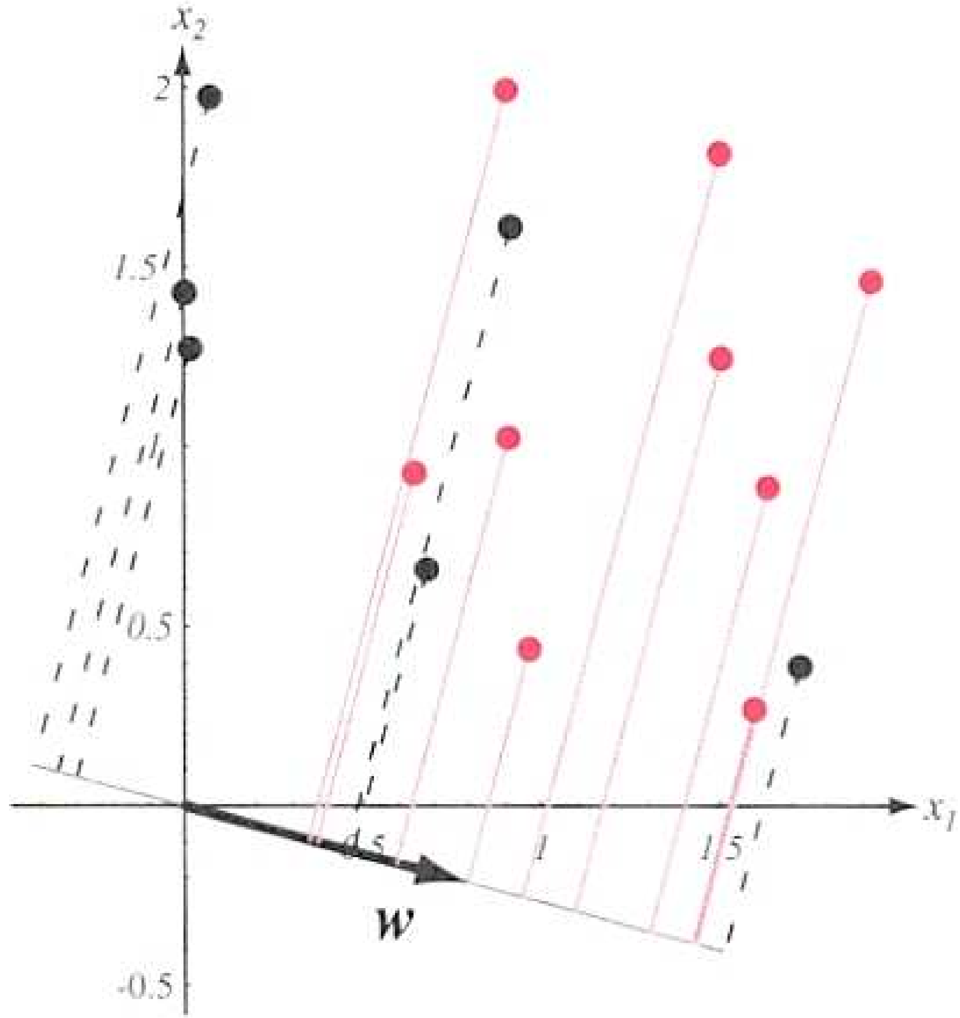


# FISHEROVA DISKRIMINACE

- ☑ 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. $\Sigma_i w_i^2=1$
- Langrangův součinitel (multiplikátor) 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 (♥) 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

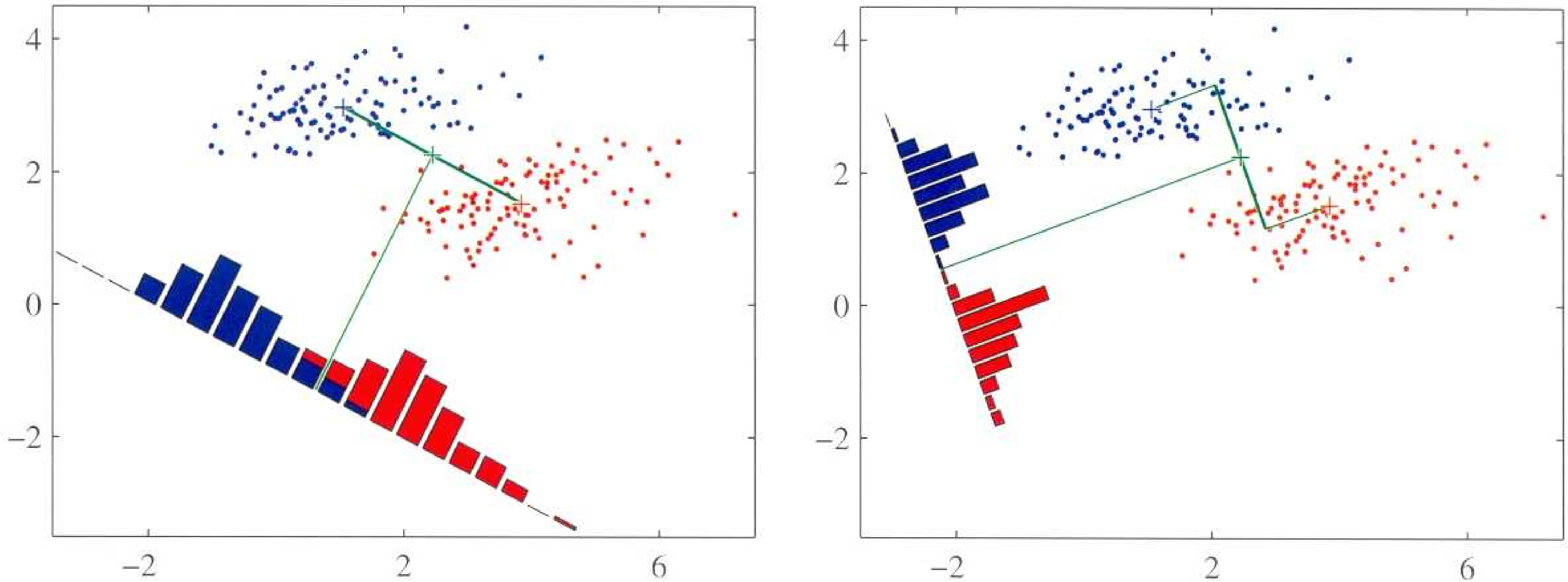
Obdobně se řeší úloha najít vázané extrémy funkce několika proměnných, např. nutná podmínka k existenci lokálního extrému funkce $w=f(x,y,z,u,v)$ při podmínkách $F_1(x,y,z,u,v)$, $F_2(x,y,z,u,v)$ je splnění rovnic

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

- ☑ variance transformovaných dat ze třídy $\omega_1$ je dána $$s_r^2 = \sum_{i \in \omega_1} (y_i - m_i)^2$$
  kde $y_i = \mathbf{w}^T \mathbf{x}_i$;
- ☑ celková variance 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:

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

kde $\mathbf{S}_B$ matice kovariance 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

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

**Fisherův diskriminátor**

  α je Langrangův multiplikátor

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 $\mathbf{m}_2$-$\mathbf{m}_1$

# FISHEROVA DISKRIMINACE
## VÍCE KLASIFIKAČNÍCH TŘÍD

předpoklady:

- počet tříd: R>2
- rozměr dat: n>R

zavedeme n′ > 1 lineárních funkcí $y_k=\mathbf{w}_k^T\mathbf{x}$, kde k=1,2,…,n′. Hodnoty $y_k$ tvoří vektor **y**. Podobně váhové vektory $\mathbf{w}_k$ reprezentují sloupce matice **W**

$$\mathbf{y} = \mathbf{W}^T\mathbf{x}$$

zobecnění matice kovariance uvnitř tříd

$$\mathbf{S}_W = \sum_{r=1}^{R} \mathbf{S}_r$$

kde

$$\mathbf{S}_r = \sum_{i \in \omega_r} (\mathbf{x}_i - \mathbf{m}_r)(\mathbf{x}_i - \mathbf{m}_r)^T \text{ a } \mathbf{m}_r = \frac{1}{n_r}.\sum_{i \in \omega_r} \mathbf{x}_i$$

$n_r$ je počet vzorků v r-té třídě

# FISHEROVA DISKRIMINACE
## VÍCE KLASIFIKAČNÍCH TŘÍD

- abychom byli schopni určit zobecněnou matici kovariance mezi třídami, určíme nejdříve celkovou kovarianční matici

$$\mathbf{S}_T = \sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T$$

- kde **m** je průměr celé množiny obrazů

$$\mathbf{m} = \frac{1}{n}.\sum_{i=1}^{n} \mathbf{x}_i = \frac{1}{n}.\sum_{r=1}^{R} n_r \mathbf{x}_i \text{ a } n = \sum_r n_r$$

- pro celkovou kovarianční matici platí $\mathbf{S}_T = \mathbf{S}_W + \mathbf{S}_B$, kde

$$\mathbf{S}_B = \sum_{r=1}^{R} n_r (\mathbf{m}_r - \mathbf{m})(\mathbf{m}_r - \mathbf{m})^T$$

# FISHEROVA DISKRIMINACE
## VÍCE KLASIFIKAČNÍCH TŘÍD

- podobně mohou být definovány matice v transformovaném n′–rozměrném prostoru **y**

$$\mathbf{s}_W = \sum_{r=1}^{R} \sum_{i \in \omega_r} (\mathbf{y}_i - \boldsymbol{\mu}_r)(\mathbf{y}_i - \boldsymbol{\mu}_r)^T \text{ a}$$

$$\mathbf{s}_B = \sum_{r=1}^{R} n_r (\boldsymbol{\mu}_r - \boldsymbol{\mu})(\boldsymbol{\mu}_r - \boldsymbol{\mu})^T,$$

kde

$$\boldsymbol{\mu}_r = \frac{1}{n_r} \sum_{i \in \omega_r} \mathbf{y}_i \text{ a } \boldsymbol{\mu} = \frac{1}{n} \sum_{r=1}^{R} n_r \boldsymbol{\mu}_r$$

# FISHEROVA DISKRIMINACE
## VÍCE KLASIFIKAČNÍCH TŘÍD

- opět chceme určit co největší skalár když je velká kovariance mezi třídami a malá kovariance uvnitř tříd

  z mnoha kritérií, např.

$$J(\mathbf{W}) = Tr\{\mathbf{s}_W^{-1}.\mathbf{s}_B\}$$

  toto kritérium může být přepsáno do tvaru

$$J(\mathbf{W}) = Tr\{(\mathbf{W}\mathbf{S}_W\mathbf{W}^T)^{-1}(\mathbf{W}\mathbf{S}_W\mathbf{W}^T)\}$$

- váhové vektory jsou v tom případě dány vlastními vektory matice $\mathbf{S}_W{}^{-1}\mathbf{S}_B$, které odpovídají jejím n′ největším vlastním číslům

Příprava nových učebních materiálů
pro obor Matematická biologie

je podporována projektem ESF
č. CZ.1.07/2.2.00/07.0318

**„VÍCEOBOROVÁ INOVACE STUDIA MATEMATICKÉ BIOLOGIE"**

EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ