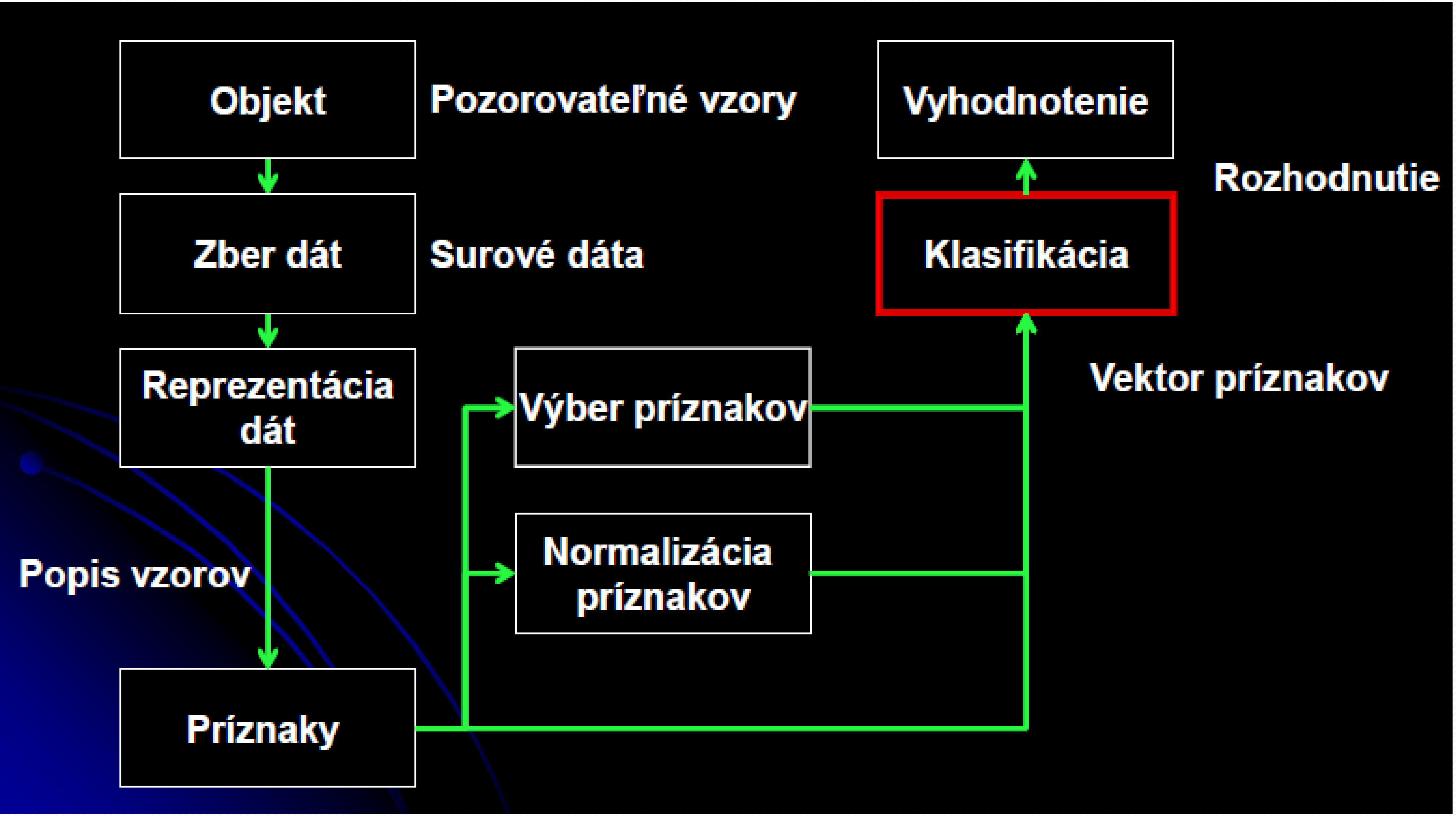
Rozpoznávanie obrazcov
šk.r. 2019-20

Lineárny klasifikátor a SVM

Doc. RNDr. Milan Ftáčnik, CSc.

Rozpoznávanie



Klasifikácia

- Rozpoznávanie obrazcov skúma metódy, ako rozhodnúť o priradení neznámeho objektu do triedy objektov – ten proces sa volá klasifikácia a zariadenie, ktoré ho robí sa volá klasifikátor
- Príznakové metódy – rozhodnutie sa robí na základe obrazca (príznakového vektora) popisujúceho objekt (najčastejšie v obraze)
- Iné metódy – priamo z obrazu objektu

Klasifikácia II

- **Príznakové metódy**
  - Diskriminačná analýza
  - Štatistické metódy
  - Vzdialenosti medzi príznakmi
  - Pravidlá (aj nemetrické príznaky)
  - Biologicky inšpirované metódy
- **Iné metódy** – konvolučné neurónové siete

Klasifikácia III

- Klasifikátor nastavujeme (trénujeme) pomocou trénovacej množiny – nastavujú sa vnútorné parametre klasifikátora
- Tento proces sa nazýva učenie a jeho cieľom je čo najmenšia chyba na trénovacej množine
- Klasifikátor môže mať aj hyperparametre, ktoré nastavuje pred učením používateľ a tie sa testujú na validačnej množine

Klasifikácia IV

- Validačná množina je časť trénovacej množiny, obyčajne v pomere trénovacia:validačná 80:20
- Kvalita klasifikátora sa testuje na testovacej množine, v ktorej sú obrazce, ktoré klasifikátor nemal k dispozícii
- Chyba dosiahnutá na testovacej množine, sa nazýva generalizačná chyba a cieľom je, aby bola čo najmenšia

Riadené a neriadené metódy

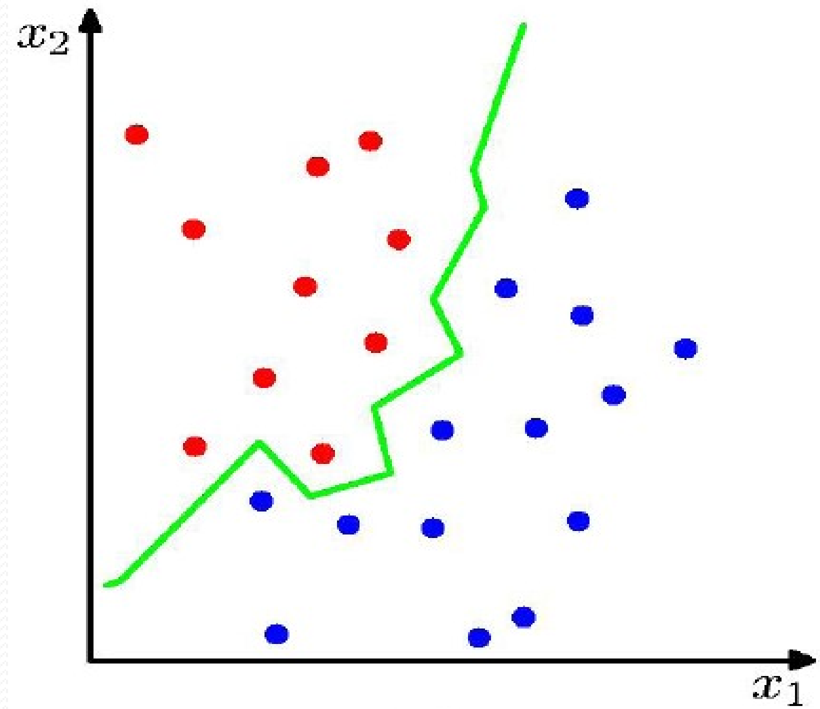
- **Riadené metódy** (učenie s učiteľom) – majú informáciu o tom, do ktorej triedy patrí každý príznakový vektor z trénovacej množiny
- Jedna premenná, ktorá sa nazýva cieľová (klasifikačná trieda), riadi proces učenia tak, aby ostatné premenné (prediktory) čo najlepšie predpovedali (na trénovacej množine známu) hodnotu cieľovej premennej

Riadené a neriadené metódy II

- **Neriadené metódy** (učenie bez učiteľa) – informáciu o zaradení do tried nemajú
- Snažia sa:
  - zhlukovať dáta do skupín podobných vzoriek
  - určiť odhad hustoty pravdepodobnosti dát
  - naučiť sa vytvárať vzorky z rozdelenia
  - nájsť krivku, ku ktorej sú dáta blízko

Riadené metódy klasifikácie

- V trénovacej množine máme *N* pozorovaní $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N), \boldsymbol{x}_i \in \mathbb{R}^D$ a vektor označenia tried $\boldsymbol{y} = [y_1 y_2 \quad y_N], y_i \in \{-1,1\}$
- **Klasifikačný problém:** Určiť $f(\boldsymbol{x})$ tak, aby $f(\boldsymbol{x}_i) = y_i$
- Niekedy sa $f(\boldsymbol{x})$ označuje ako rozhodovacia funkcia



Diskriminačná analýza

- Predpokladáme, že klasifikačné triedy sú **lineárne separovateľné**
- To znamená, že pre každé dve triedy existuje nadrovina rozdeľujúca príznakový priestor tak, že príznakové vektory (z trénovacej množiny) z rôznych tried ležia v rôznych polpriestoroch
- Nadrovina je $D - 1$ rozmerný podpriestor, ktorý delí príznakový priestor na dve časti

Diskriminačná analýza II

- Nájdeme lineárne diskriminačné funkcie, $g_1(\boldsymbol{x}), \dots, g_R(\boldsymbol{x})$, ktorých bude toľko čo tried
- Lineárna diskriminačná funkcia má tvar:

$$g_r(\boldsymbol{x}) = q_{r0} + q_{r1}x_1 + \dots + q_{rn}x_n$$

kde q_{ri} sú skalárne koeficienty

- Neznámy vektor $\boldsymbol{x}$ zaradíme do tej triedy, ktorej diskriminačná funkcia vrátila najväčšiu hodnotu

Diskriminačná analýza III

- Z toho vyplýva rozhodovacie pravidlo

$$f(\boldsymbol{x}) = \omega_r \leftrightarrow g_r(\boldsymbol{x}) = \max_{s=1,2,\ldots,R} g_s(\boldsymbol{x})$$

- Potom je nadrovina medzi dvoma triedami ω_r a ω_s určená vzťahom

$$g_r(\boldsymbol{x}) - g_s(\boldsymbol{x}) = 0$$

- Pre dve triedy nám stačí jedna diskriminačná funkcia; ak bude kladná, zaradíme vektor do ω_1 a ak bude záporná, tak do ω_2

Klasifikácia do dvoch tried

- Definujme funkciu $f(\boldsymbol{x}) = g_1(\boldsymbol{x}) - g_2(\boldsymbol{x})$
- Ak označíme rozdiel skalárov pri jednotlivých príznakoch x_i ako vektor $\boldsymbol{w}$ a skalár q_0 ako b, tak môžeme napísať vektorovo

$$f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b$$

- Potom rovnica nadroviny je $\boldsymbol{w}^T\boldsymbol{x} + b = 0$ kde $\boldsymbol{w}$ predstavuje váhový vektor a platí

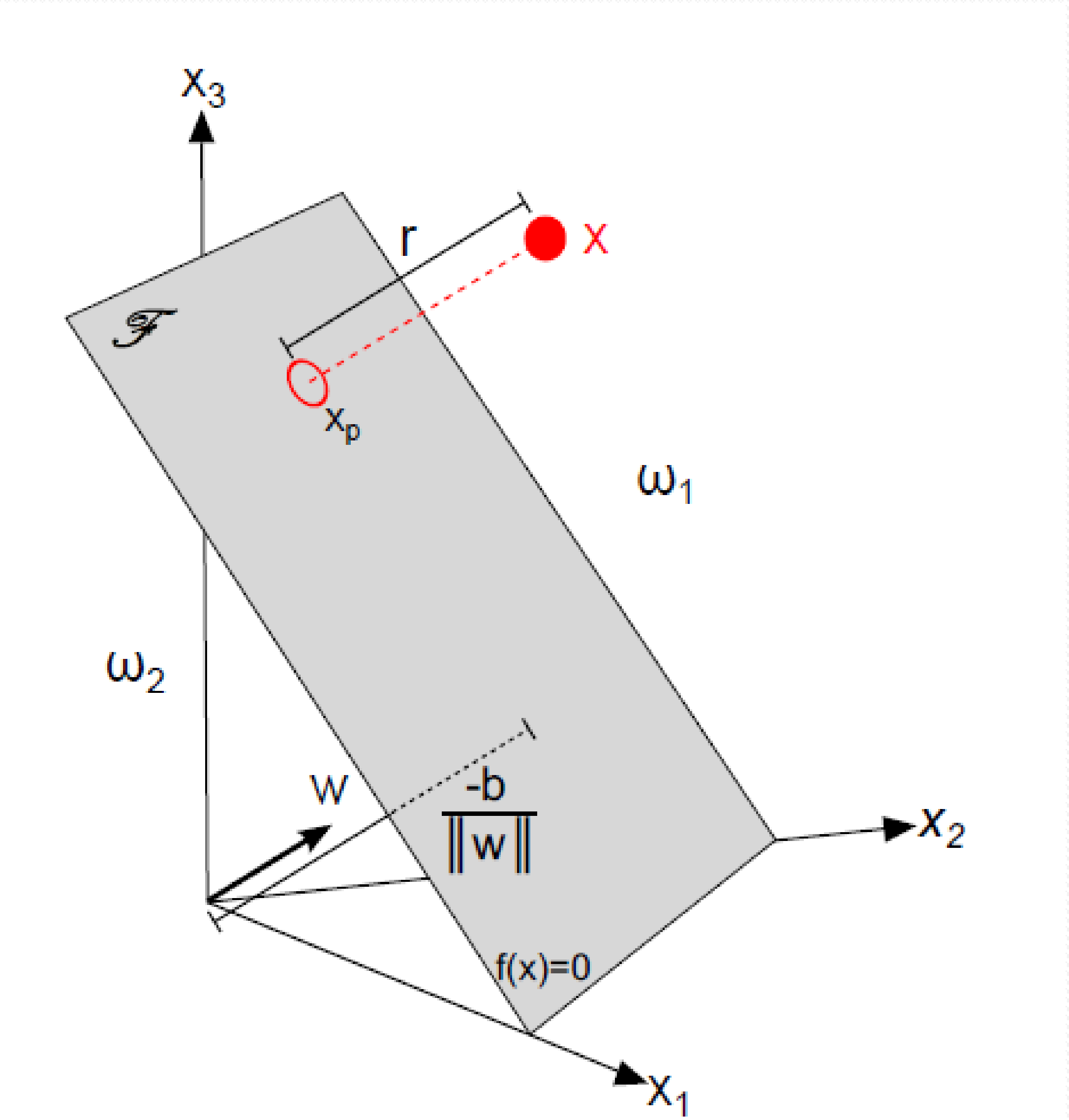
$$f(\boldsymbol{x}) > 0 \text{ ak } \boldsymbol{x} \in \omega_1$$

$$f(\boldsymbol{x}) < 0 \text{ ak } \boldsymbol{x} \in \omega_2$$

Klasifikácia do dvoch tried II

- Ak budeme uvažovať vektory $\boldsymbol{x}_1$ a $\boldsymbol{x}_2$, ktoré ležia na oddeľujúcej nadrovine H, tak platí $\boldsymbol{w}^T\boldsymbol{x}_1 + b = \boldsymbol{w}^T\boldsymbol{x}_2 + b$ alebo $\boldsymbol{w}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2) = 0$
- Z toho vyplýva, že $\boldsymbol{w}$ je normála ku každému vektoru, ktorý leží v H
- Pretože $f(\boldsymbol{x}) > 0$ pre všetky vektory z triedy ω_1, tak normálový vektor $\boldsymbol{w}$ ukazuje do pol-priestoru zodpovedajúcemu ω_1, a tiež hovoríme, že je na pozitívnej strane H

Klasifikácia do dvoch tried III



Klasifikácia do dvoch tried IV

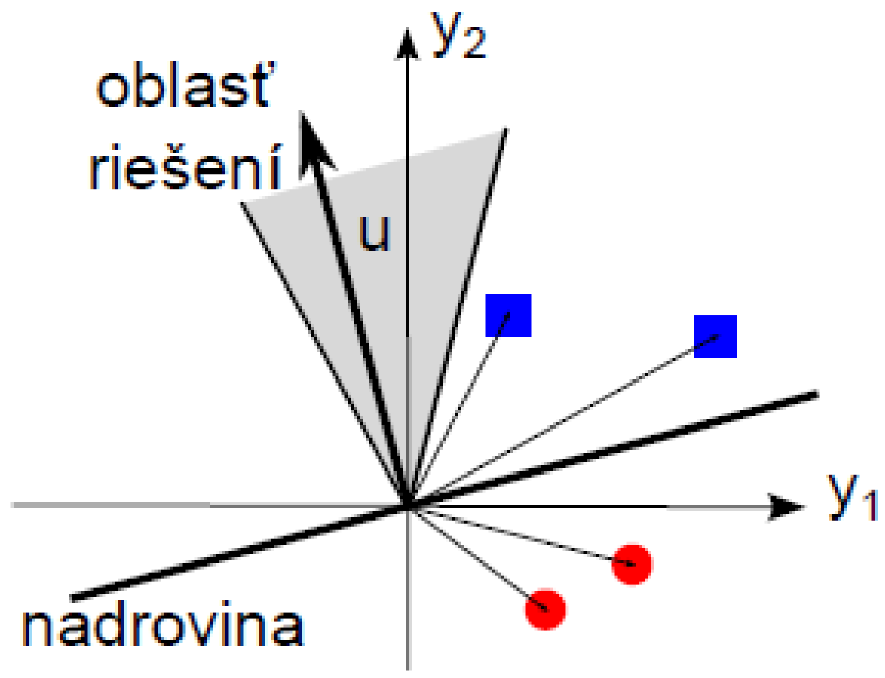
- Diskriminačná funkcia $f(\boldsymbol{x})$ je algebraickou mierou vzdialenosti $\boldsymbol{x}$ od H
- Vidíme to, ak si označíme priemet $\boldsymbol{x}$ na H ako $\boldsymbol{x}_p$, potom môžeme napísať $\boldsymbol{x} = \boldsymbol{x}_p + r\frac{\boldsymbol{w}}{||\boldsymbol{w}||}$
- Pretože platí $f(\boldsymbol{x}_p) = \boldsymbol{w}^T\boldsymbol{x}_p + b = 0$, z toho úpravou a s využitím vzťahu $\boldsymbol{w}^T\boldsymbol{w} = ||\boldsymbol{w}||^2$ dostaneme $$r = \frac{f(\boldsymbol{x})}{||\boldsymbol{w}||}$$

Klasifikácia do dvoch tried V

- Ako dôsledok z toho vyplýva, že vzdialenosť od počiatku k H je $\frac{b}{||\boldsymbol{w}||}$
- Ak je $b > 0$, tak počiatok leží na pozitívnej strane H, ak je $b < 0$, tak na negatívnej
- Ak je $b = 0$, tak oddeľujúca nadrovina H prechádza cez počiatok
- Orientácia nadroviny H je teda daná normálovým vektorom $\boldsymbol{w}$ a jej poloha členom b

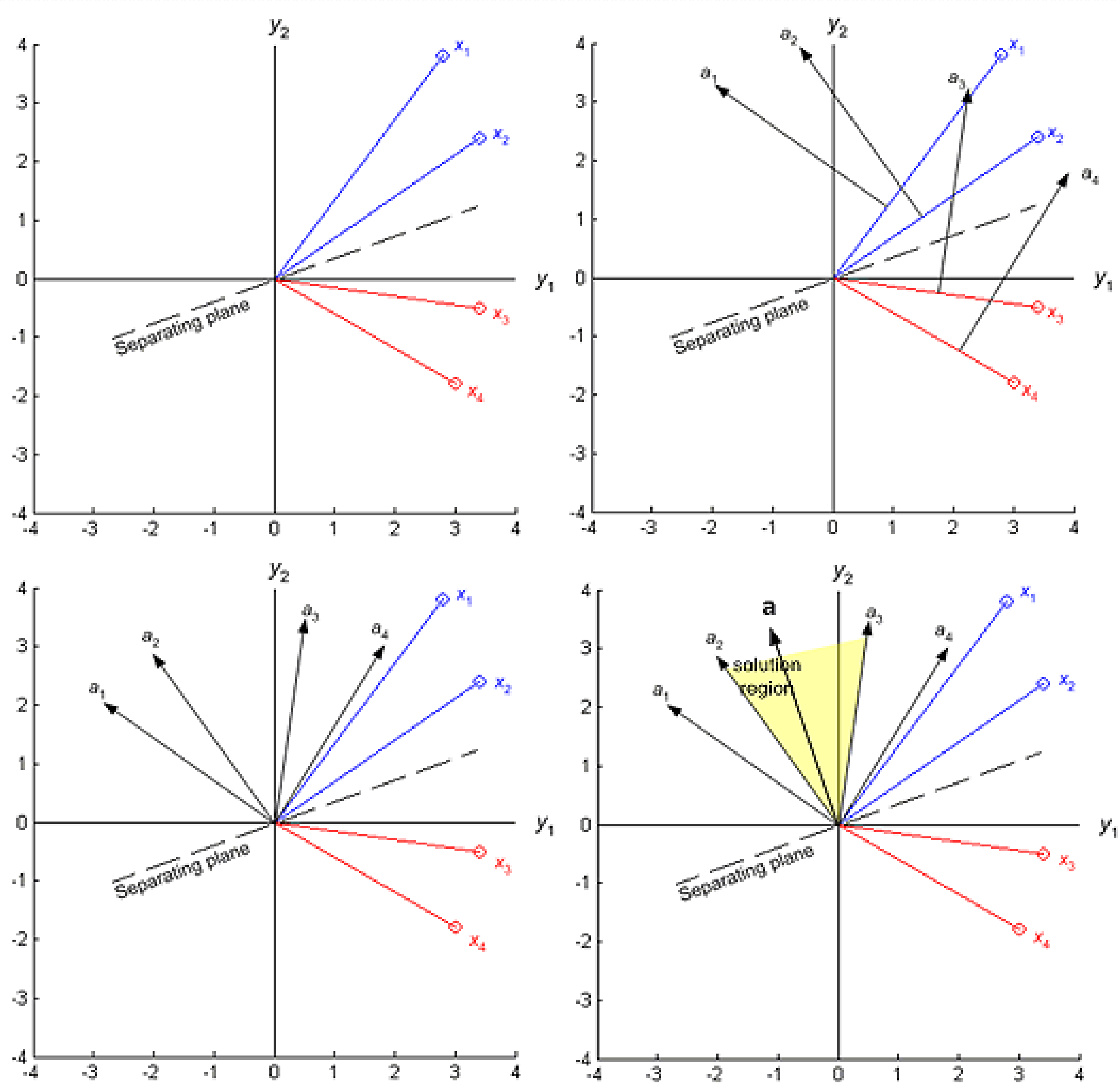
Zovšeobecnený lineárny klasifikátor

- Jeho funkcia je $f(\boldsymbol{x}) = \boldsymbol{u}^T\boldsymbol{y}$
  kde $\boldsymbol{u} = [b, \boldsymbol{w}^T]^T$ a $\boldsymbol{y} = [1, \boldsymbol{x}^T]^T$
- Vzorku $\boldsymbol{y}_i$ zaradíme
  - do triedy ω_1 ak platí $\boldsymbol{u}^T\boldsymbol{y}_i > 0$
  - do triedy ω_2 ak platí $\boldsymbol{u}^T\boldsymbol{y}_i < 0$



Zovšeobecnený lineárny klasifikátor II

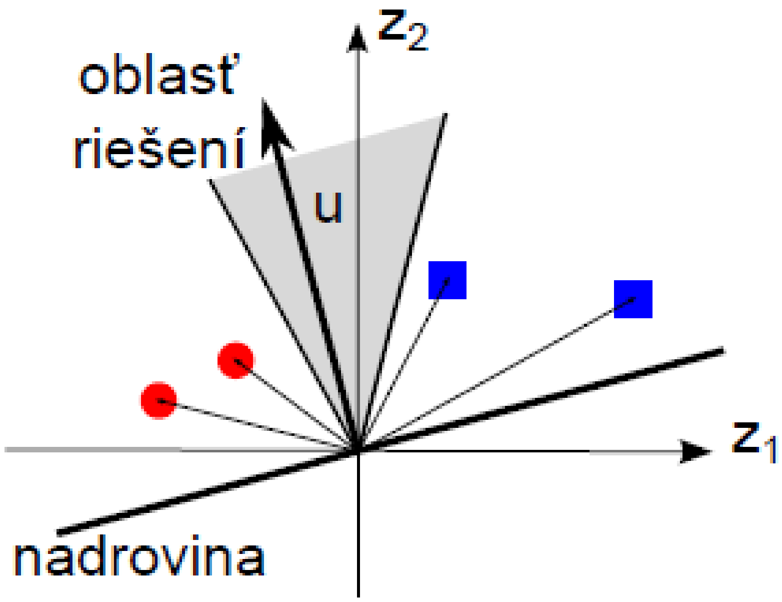
- Oblasť riešení sú možné polohy vektora $\boldsymbol{u}$ (na obrázku označe-ného ako $\boldsymbol{a}$)



Zovšeobecnený lineárny klasifikátor III

- Označíme

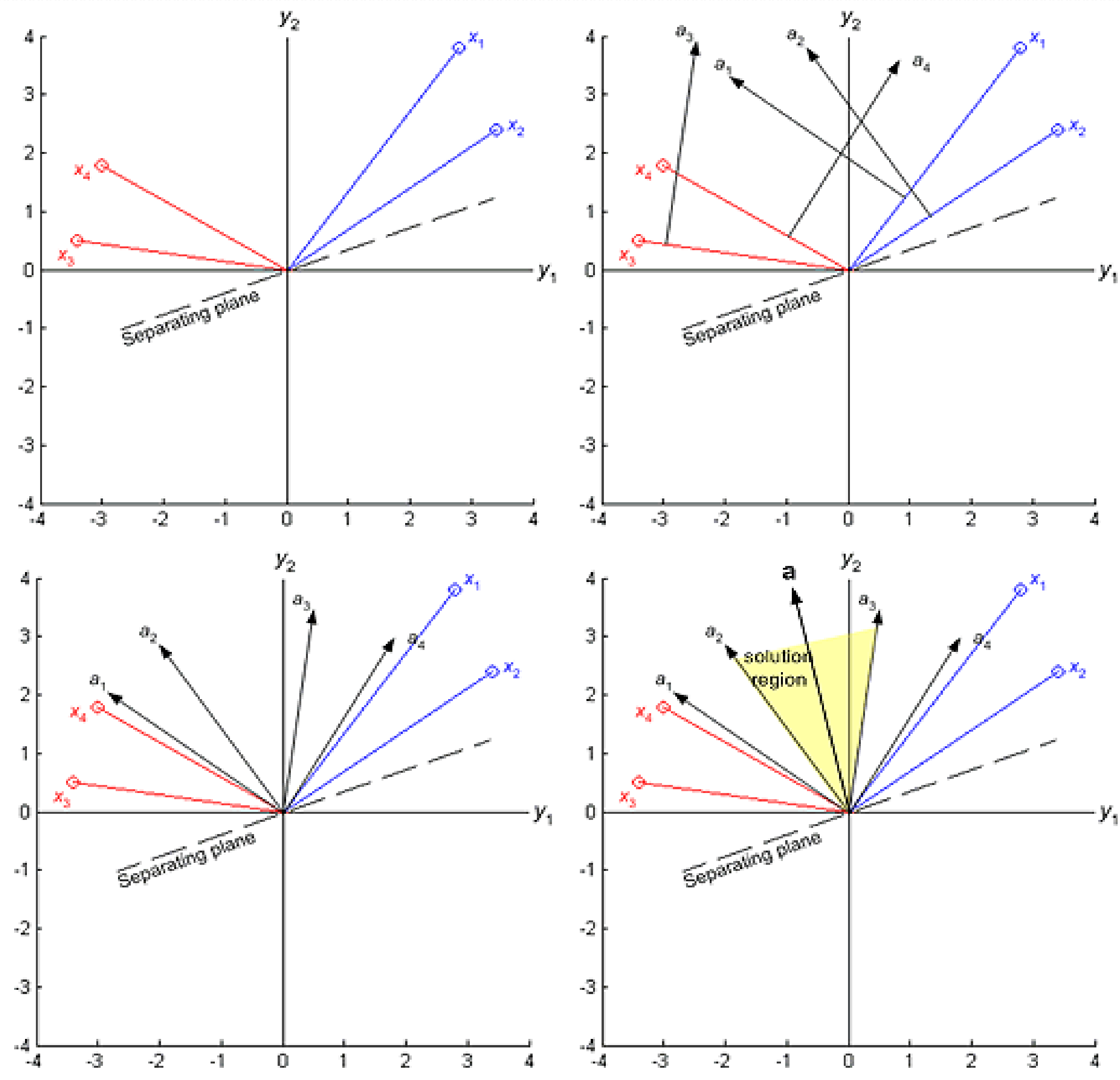
$$\mathbf{z}_i = \begin{cases} \mathbf{y}_i & \text{pre } \mathbf{y} \in \omega_1, \\ -\mathbf{y}_i & \text{pre } \mathbf{y} \in \omega_2, \end{cases}$$



- Našou úlohou je nájsť vektor $\boldsymbol{u}$, aby platil pre všetky $\boldsymbol{z}_i$ vzťah $\boldsymbol{u}^T\boldsymbol{z}_i > 0$

Zovšeobecnený lineárny klasifikátor IV

- Oblasť riešení sú možné polohy vektora $\boldsymbol{u}$ (na obrázku označeného ako $\boldsymbol{a}$)



Gradientná metóda

- Namiesto riešenia lineárnych nerovníc zvolíme skalárnu účelovú funkciu $O(\boldsymbol{u})$, ktorej hodnota je minimálna, keď $\boldsymbol{u}$ je riešením
- Začneme náhodne zvoleným vektorom u_1
- Vypočítame gradient ∇O v bode u_1
- Aktualizujeme hodnotu u tak, že vyberieme vektor z okolia $\boldsymbol{u}_1$ v smere najväčšej negatívnej zmeny gradientu $\boldsymbol{u}_{i+1} = \boldsymbol{u}_i - \eta(i)\ \nabla O|_{\boldsymbol{u}_i}$,

kde $\eta(i)$ je (pozitívna) rýchlosť učenia v kroku *i*

Gradientná metóda II

- Úspešnosť tejto metódy závisí od voľby $\eta(i)$
- Využijeme Taylorov rozvoj a to, že účelová funkcia je diferencovateľná
- $O(\boldsymbol{a})$ je v okolí bodu $\boldsymbol{u}_i$ aproximovaná

$$O(\mathbf{a}) \approx O(\mathbf{u}_i) + \nabla O^T(\mathbf{a}-\mathbf{u}_i) + \frac{1}{2}(\mathbf{a}-\mathbf{u}_i)^T\mathbf{H}(\mathbf{a}-\mathbf{u}_i)$$

kde $\mathbf{H}$ je matica druhých parciálnych derivácií funkcie O (Hesseho matica)

Gradientná metóda III

- Ak za $\boldsymbol{a}$ dosadíme $\boldsymbol{u}_{i+1}$, tak dostaneme

$$O(\mathbf{u}_{i+1}) \approx O(\mathbf{u}_i) - \eta(i)\|\nabla O\|^2 + \frac{1}{2}\eta^2(i)\nabla O^T \mathbf{H} \nabla O$$

- Dá sa ukázať, že minimum nadobúda táto funkcia pre

$$\eta(i) = \frac{\|\nabla O\|^2}{\nabla O^T \mathbf{H} \nabla O}.$$

- Často je rýchlejšie určiť $\eta(i)$ ako konštantu η menšiu ako je potrebné a urobiť o niečo viac krokov ako v každom počítať optimálne $\eta(i)$

Newtonova metóda

- Je založená na funkcii $\boldsymbol{u}_{i+1} = \boldsymbol{u}_i - \boldsymbol{H}^{-1}\nabla O$ a hľadá riešenie až kým $\boldsymbol{H}^{-1}\nabla O < \Theta$, kde Θ je prahová hodnota veľkosti chyby
- Newtonov algoritmus dáva väčšinou v jednom kroku lepšie zlepšenie ako gradientná metóda, dokonca aj s optimálnou hodnotou $\eta(i)$
- Ale Newtonov algoritmus sa nedá použiť, ak je Hessovská matica $\boldsymbol{H}$ singulárna

Perceptrónová kriteriálna funkcia

- Využíva iba nesprávne klasifikované vzorky
- Účelová funkcia je po častiach lineárna perceptrónová funkcia

$$O_P = \sum_{\mathbf{z} \in \mathcal{Z}} (-\mathbf{u}^T \mathbf{z})$$

kde $\mathcal{Z}$ je množina nesprávne klasifikovaných vzoriek. Potom $\nabla O_P = \sum_{\mathbf{z} \in \mathcal{Z}} (-\mathbf{z})$ a

$$\mathbf{u}_{i+1} = \mathbf{u}_i + \eta(i) \sum_{\mathbf{z} \in \mathcal{Z}_i} \mathbf{z}$$

Perceptrónová funkcia II

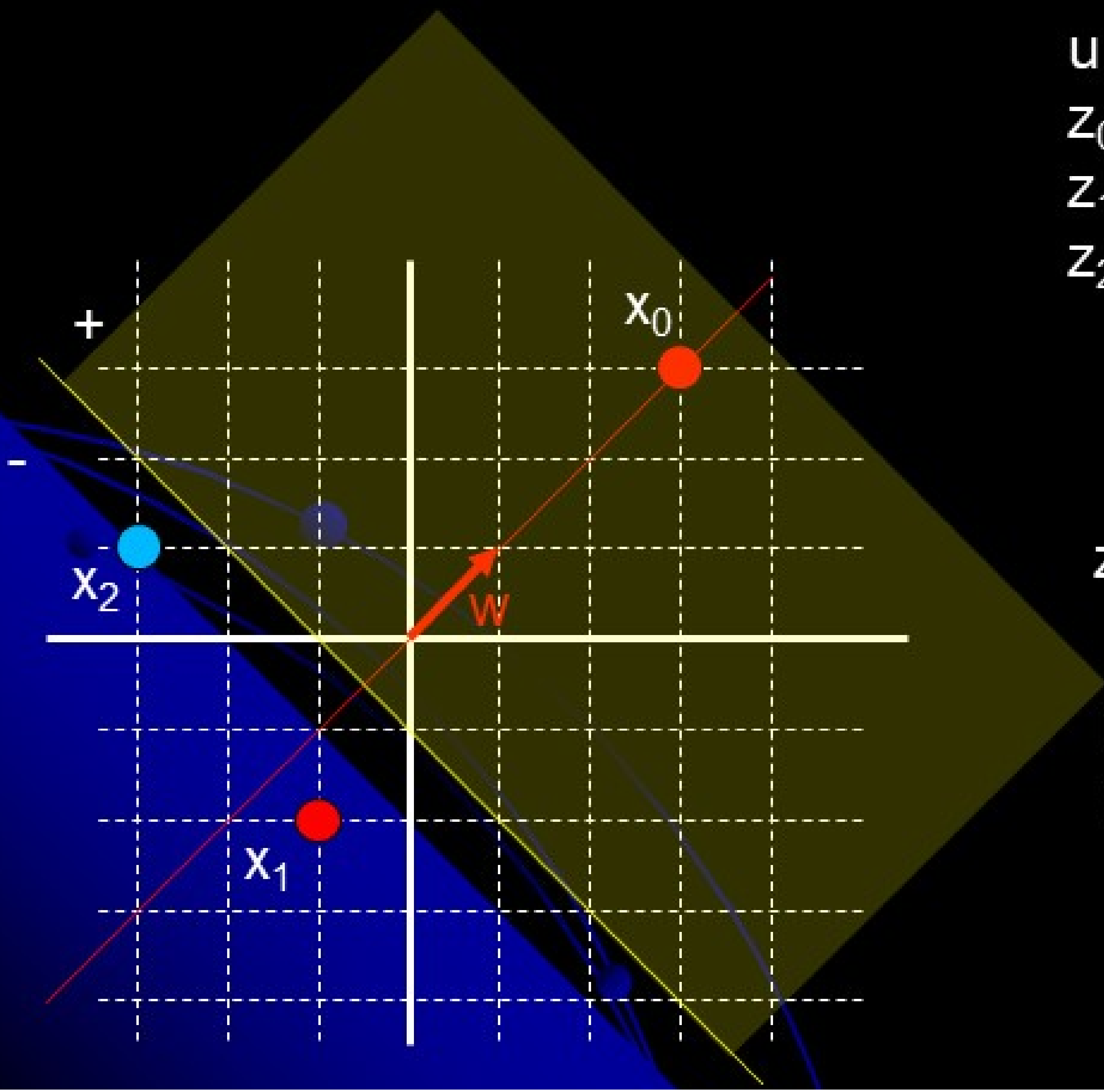
- V prípade, že ak je množina O_P prázdna, tak definujeme $O_P \equiv 0$
- Vieme, že pre nesprávne klasifikované vzorky platí $\boldsymbol{u}^T\boldsymbol{z} \leq 0$, a preto O_P nadobúda nezáporné hodnoty
- Nulovú hodnotu nadobúda pre váhové vektory, ktoré sú riešením nášho problému, keď nemáme nesprávne klasifikované vzorky

Perceptrónová funkcia III

- Tento prístup sa volá dávkový, pretože vyhodnocuje všetky zle klasifikované vzorky naraz
- Postup pri vyhodnocovaní vzoriek po jednej

```
inicializuj u_0 = 1, i=0, k=0
opakuj
   ak z_k je nesprávne klasifikovaný
      u_{i+1} = u_i + z_k
      i=i+1
   koniec ak
   k=(k+1) mod N
kým existuje nejaký nesprávne klasifikovaný vektor z_j
```

Ukážka postupu algoritmu



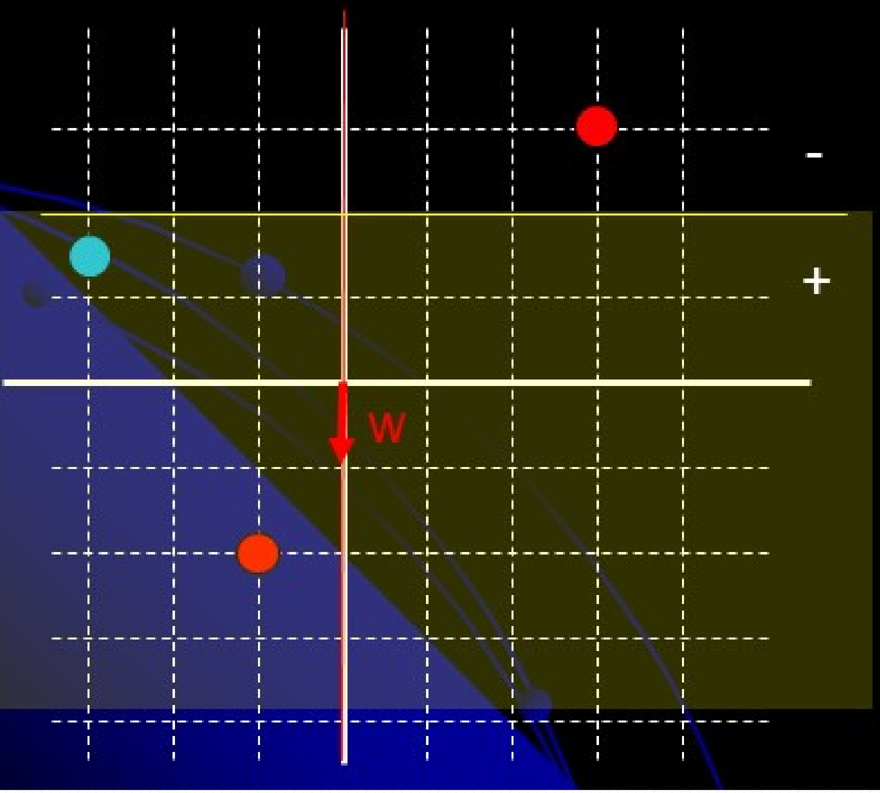
$u = (1,1,1)$
$z_0 = (1,3,3)$
$z_1 = (1,-1,-2)$
$z_2 = -(1,-3,1)$

z_1 zle klasifikovaný

$u = u + z_1 = (1,1,1) + (1,-1,-2)$
$(2,0,-1)$

Ukážka postupu algoritmu II

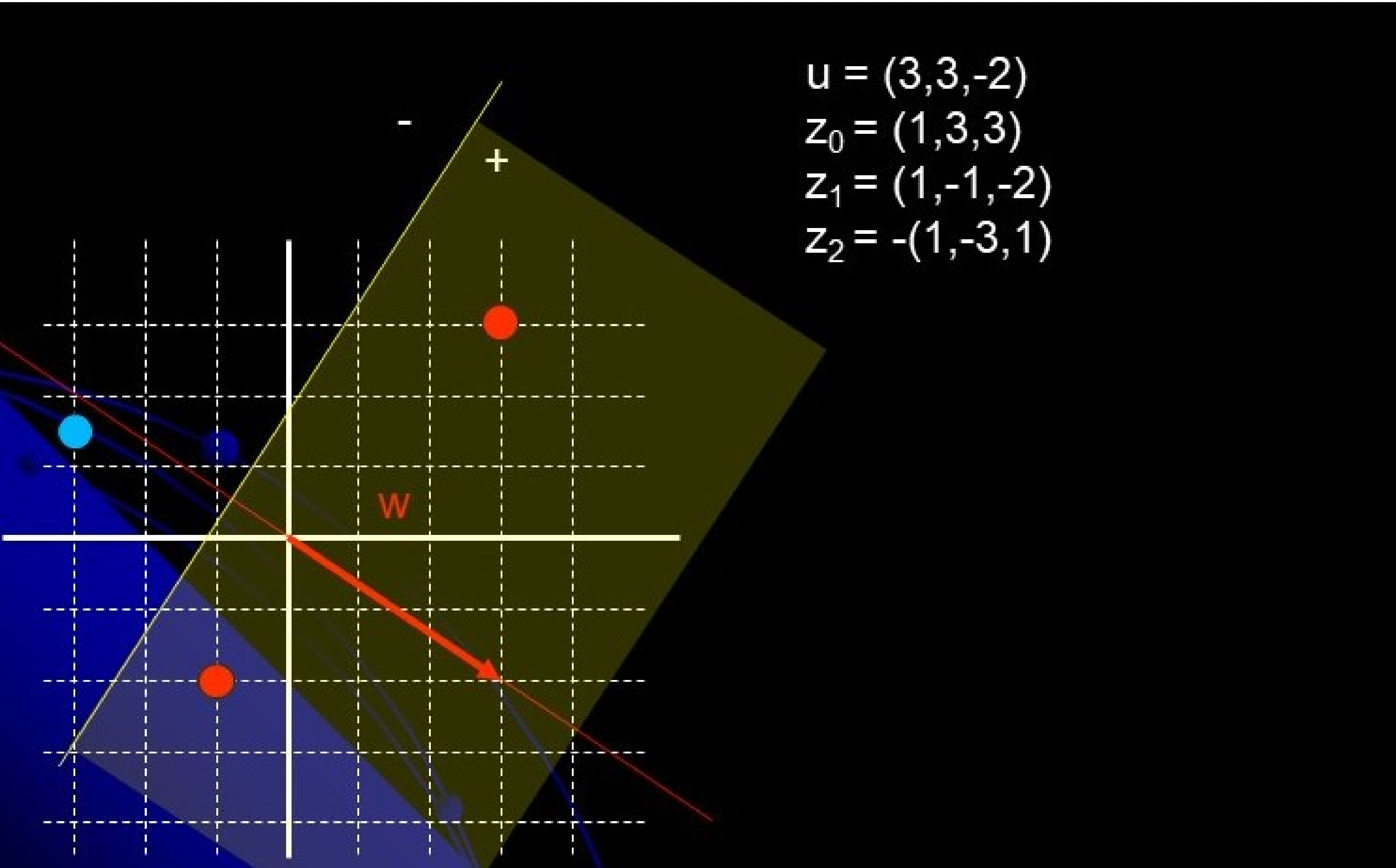
$u = (2,0,-1)$
$z_0 = (1,3,3)$
$z_1 = (1,-1,-2)$
$z_2 = -(1,-3,1)$



z_2 zle klasifikovaný

$u = u + z_2 = (2,0,-1) - (1,-3,1) = (3,3,-2)$

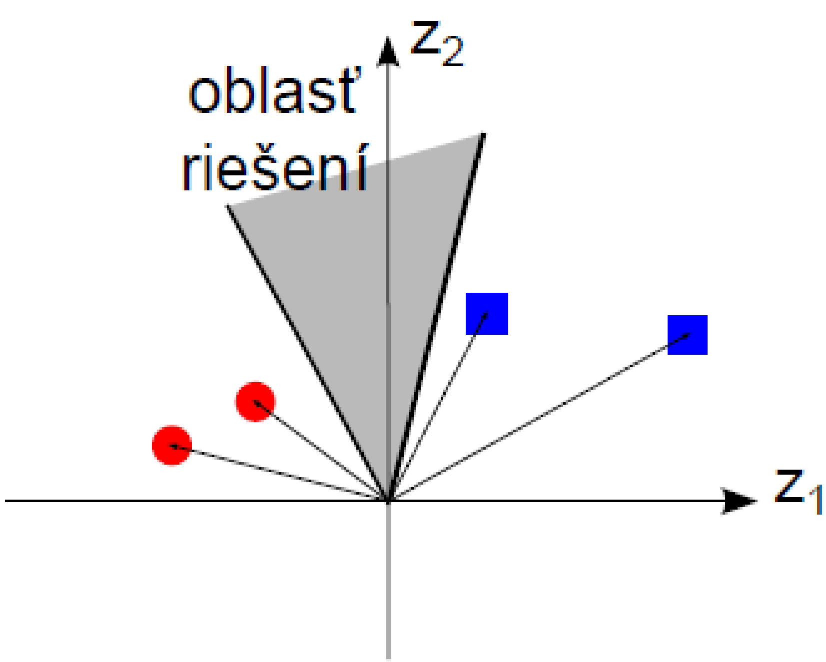
Ukážka postupu algoritmu III



Kvadratická účelová funkcia

$$O_Q = \sum_{\mathbf{z} \in \mathcal{Z}_i} (\mathbf{u}^T \mathbf{z})^2$$

- Má spojitý gradient, je príliš hladká
- Môže konvergovať na hranici oblasti riešení
- Dlhšie vektory $\boldsymbol{z}$ vplývajú na hodnotu O_Q viac ako kratšie

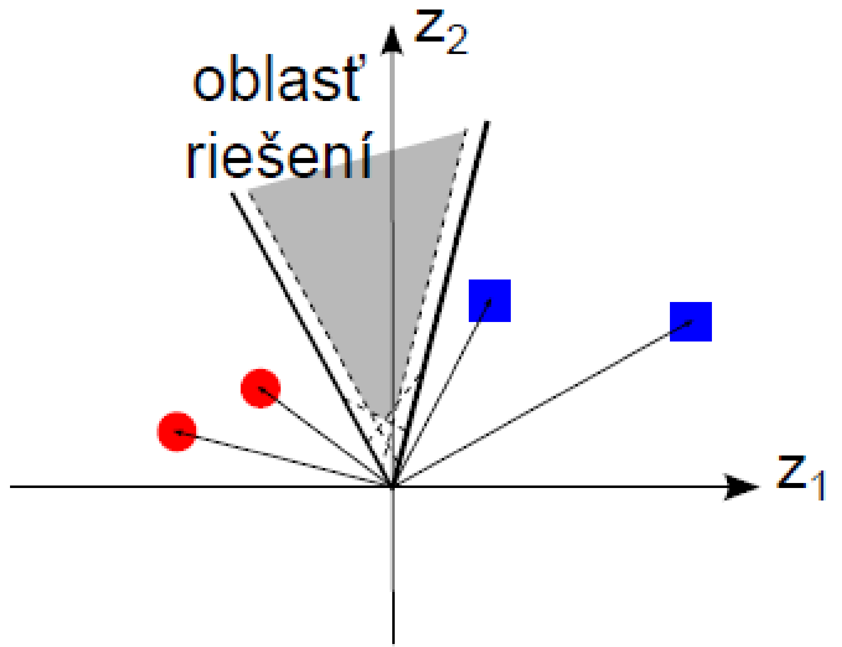


Normalizačný člen

- Pridáme člen určujúci okraj v oblasti riešení

$$O_R = \frac{1}{2} \sum_{\mathbf{z} \in \mathcal{Z}_i} \frac{(\mathbf{u}^T \mathbf{z} - \epsilon)^2}{\|\mathbf{z}\|^2}.$$

$$\nabla O_R = \sum_{\mathbf{z} \in \mathcal{Z}_i} \frac{\mathbf{u}^T \mathbf{z} - \epsilon}{\|\mathbf{z}\|^2} \mathbf{z}.$$



Úprava algoritmu

- Po doplnení normalizačného člena

```
inicializuj u_0 = 0, i=0, k=0
opakuj
    k=(k+1) mod N
    ak u^T z_k ≤ ϵ
        z = z_k
        u_{i+1} = u_i + η(i) (ϵ − u^T z)/‖z‖^2 z
        i=i+1
    koniec ak
kým u^T z_j > ϵ pre všetky z_j
```

inicializuj $\mathbf{u}_0 = 0$, i=0, k=0

opakuj

k=(k+1) mod N

ak $\mathbf{u}^T\mathbf{z}_k \leq \epsilon$

$\mathbf{z} = \mathbf{z}_k$

$\mathbf{u}_{i+1} = \mathbf{u}_i + \eta(i)\frac{\epsilon - \mathbf{u}^T\mathbf{z}}{\|\mathbf{z}\|^2}\mathbf{z}$

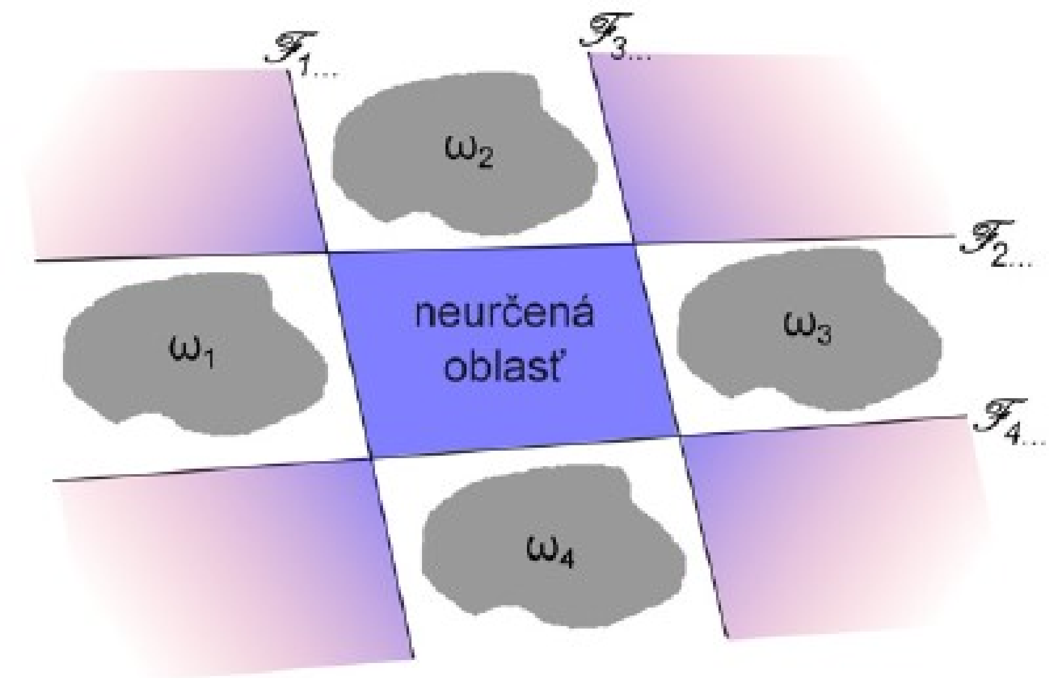
i=i+1

koniec ak

kým $\mathbf{u}^T\mathbf{z}_j > \epsilon$ pre všetky $\mathbf{z}_j$

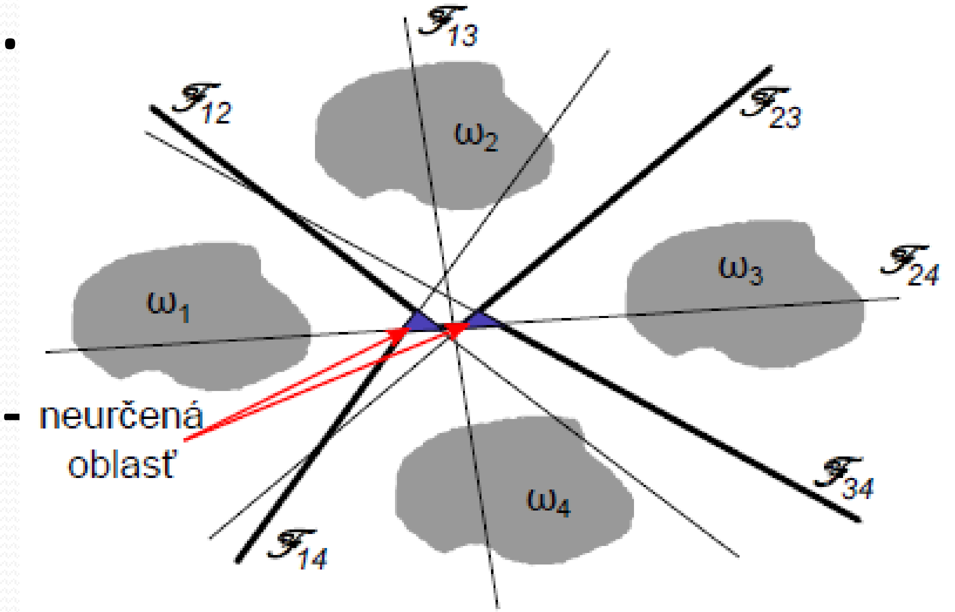
Klasifikácia do *R*-tried

- 1. Pre každdú klasifikačnú triedu môžeme vytvoriť rozhodovaciu funkciu, ktorá oddelí vzorky tejto triedy od vzoriek ostatných tried.
- Takto budeme mať *R* rozhodovacích funkcií
- Niektoré oblasti môžu byť neurčené



Klasifikácia do *R*-tried II

- 2. Pre každú dvojicu rôznych klasifikačných tried vytvoríme rozhodovaciu funkciu oddeľu-júcu vzorky týchto dvoch tried bez ohľadu na vzorky ostatných tried.
- Dostaneme $R(R-1)/2$ rozhodovacích funkcií
- Aj tu môžu byť neurče-né oblasti



Lineárny stroj

- Preto vytvoríme *R* rozhodovacích funkcií

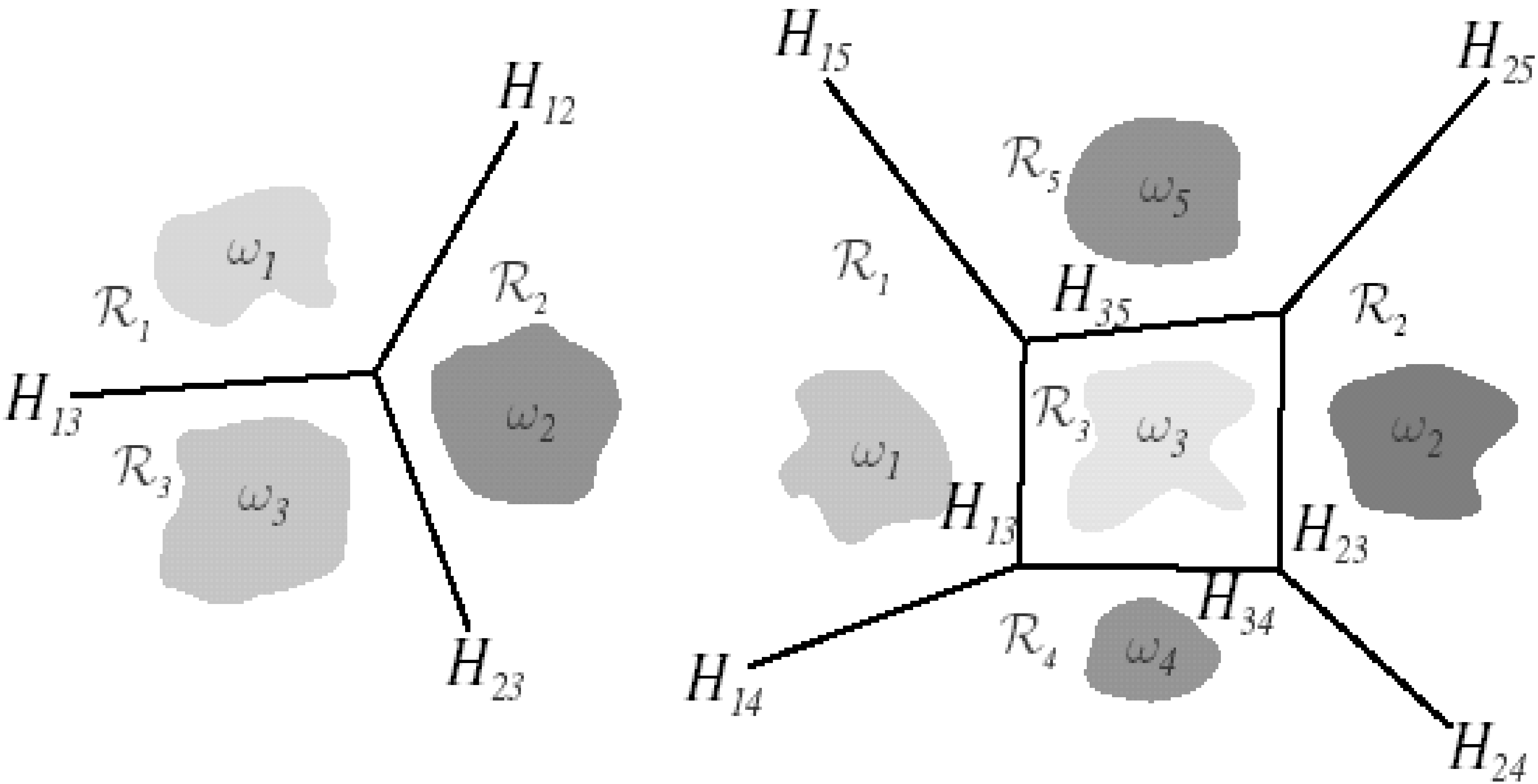
$$\{f_k(\boldsymbol{x})\}_{k=1}^{R}$$

- $\boldsymbol{x}$ zaradíme do tej klasifikačnej triedy, ktorej hodnota rozhodovacej funkcie po dosadení je najväčšia

$$\boldsymbol{x} \in \omega_i \leftrightarrow \forall_{j \neq i}\, f_i(\boldsymbol{x}) > f_j(\boldsymbol{x})$$

- To rozdelí priestor do *R* konvexných oblastí

Lineárny stroj II



Lineárny stroj III

- Pre dve susedné oblasti zodpovedajúce ω_i a ω_j je hranica medzi nimi súčasťou nadroviny H_{ij} definovanej rovnosťou

$$f_i(\boldsymbol{x}) = f_j(\boldsymbol{x}) \text{ alebo } (\boldsymbol{w}_i - \boldsymbol{w}_j)^T \boldsymbol{x} + (b_i - b_j) = 0$$

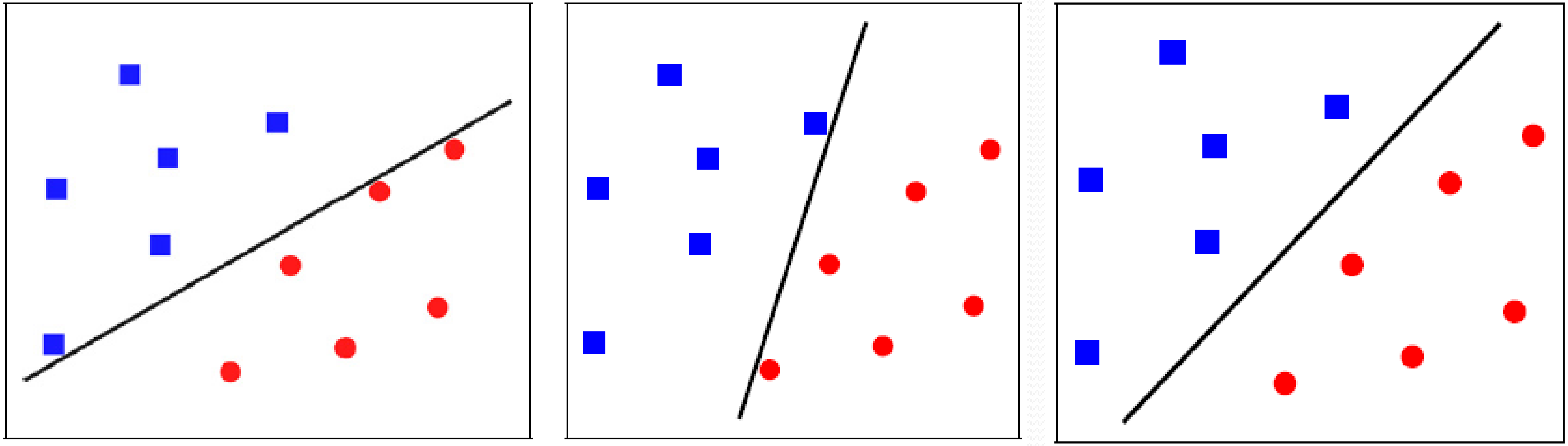
- Z toho vyplýva, že $\boldsymbol{w}_i - \boldsymbol{w}_j$ je normála ku H_{ij} a signovaná vzdialenosť od $\boldsymbol{x}$ ku H_{ij} určená vzťahom $\frac{(f_i(\boldsymbol{x}) - f_j(\boldsymbol{x}))}{\|\boldsymbol{w}_i - \boldsymbol{w}_j\|}$

Lineárny stroj IV

- Takže pre lineárny stroj to nie sú váhové vektory, ktoré sú dôležité, ale ich rozdiely
- Celkove je $R(R-1)/2$ dvojíc oblastí, ktoré zodpovedajú triedam, ale nie všetky musia byť susedné
- Potom celkový počet segmentov nadrovín, ktoré vystupujú v rozhodovacích plochách, môže byť nižší

Support vector machines (SVM)

- Máme nadrovinu, ktorá síce dobre klasifikuje trénovacie dáta, ale pri testovacích dátach dosahuje veľkú chybu
- Čiastočné riešenie: použitie účelovej funkcie O_R, ale optimálna hodnota parametra nie je

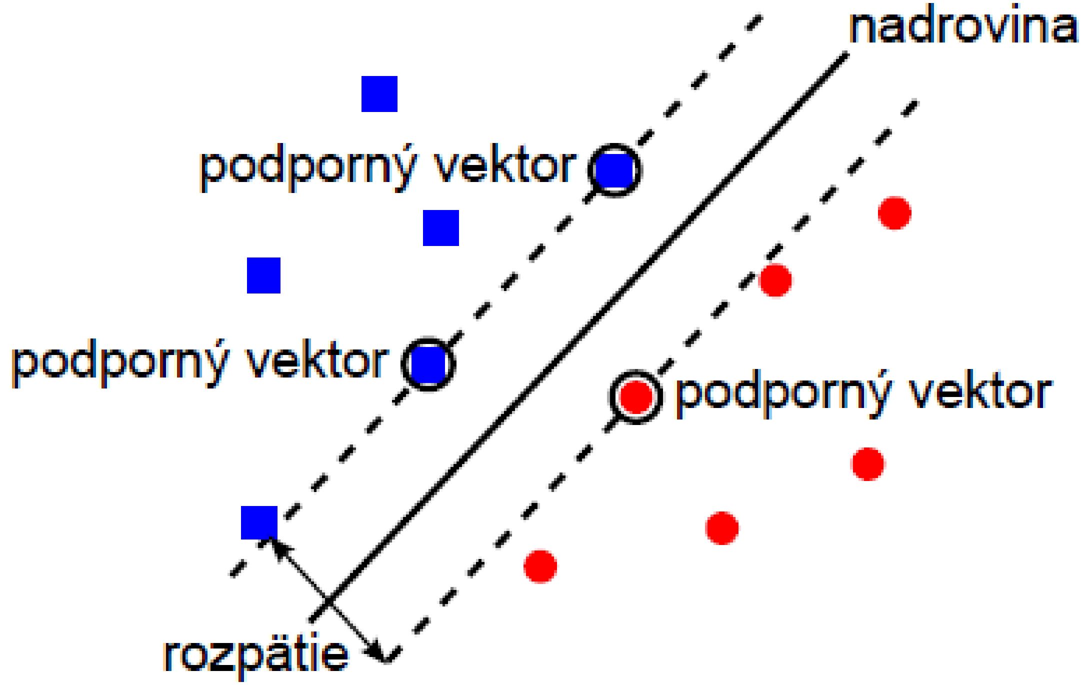


SVM II

- Mechanizmus podporných vektorov (SVM) sa snaží nájsť nadrovinu, čo najviac vzdialenú od trénovacích vzoriek dvoch tried
- Podpornými vektormi sú potom tie vzorky každej triedy, ktoré sú najbližšie k oddeľujúcej nadrovine.
- Vzdialenosť medzi podpornými vektormi rôznych tried sa nazýva rozpätie (margin)

SVM III

- Pre takúto nadrovinu platí $\boldsymbol{w}^T x + b = \pm 1$ a rozpätie $\frac{2}{||\boldsymbol{w}||}$



SVM IV

- Hľadáme maximum

$$\mathbf{w}^* = \arg\max \frac{2}{\|\mathbf{w}\|}$$

- A súčasne minimum

$$(\mathbf{w}^*, b^*) = \arg\min \frac{1}{2} \|\mathbf{w}\|^2,$$

aby

$$\mathbf{w}^T \mathbf{x}_i + b \geq 1 \text{ pre } \mathbf{x} \in \omega_1,$$

$$\mathbf{w}^T \mathbf{x}_i + b \leq -1 \text{ pre } \mathbf{x} \in \omega_2.$$

- Ako na to použiť Lagrangeovu funkciu?

SVM V

- Pomôže trik, aby podmienka bola iba jedna
- Spojíme podmienky $(\mathbf{w}^T\mathbf{x}_i + b)k_i \geq 1$ kde

$$k_i = \begin{cases} 1 & \text{pre } \mathbf{x} \in \omega_1 \\ -1 & \text{pre } \mathbf{x} \in \omega_2 \end{cases}$$

- Lagrangián: $L(\mathbf{w}, b, \alpha_i) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i((\mathbf{w}^T\mathbf{x}_i + b)k_i - 1)$

$$= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i(\mathbf{w}^T\mathbf{x}_i + b)k_i + \sum_{i=1}^{N} \alpha_i,$$

aby

$$\alpha_i \geq 0.$$

Kuhn-Tuckerove podmienky

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{N} \alpha_i k_i \mathbf{x}_i \equiv 0$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{N} \alpha_i k_i \equiv 0$$

$$(\mathbf{w}^T \mathbf{x}_i + b) k_i - 1 \geq 0$$

$$\alpha_i \geq 0$$

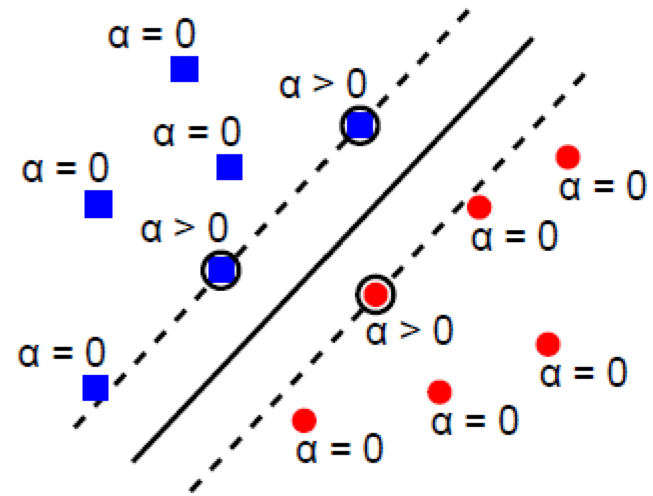
$$\alpha_i((\mathbf{w}^T \mathbf{x}_i + b) k_i - 1) = 0$$

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i k_i \mathbf{x}_i$$

$$\sum_{i=1}^{N} \alpha_i k_i = 0$$

$$\alpha_i = 0 \Leftrightarrow (\mathbf{w}^{\mathbf{T}} \mathbf{x}_i + b) k_i > 1$$

$$\alpha_i > 0 \Leftrightarrow (\mathbf{w}^{\mathbf{T}} \mathbf{x}_i + b) k_i = 1$$



Kuhn-Tuckerove podmienky II

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i k_i \mathbf{x}_i$$

$$\sum_{i=1}^{N} \alpha_i k_i = 0$$

$$L(\mathbf{w}, b, \alpha_i) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i (\mathbf{w}^{\mathbf{T}} \mathbf{x}_i + b) k_i + \sum_{i=1}^{N} \alpha_i$$

- Aby sme našli α, spočítame duálny problém

$$\boxed{\widetilde{\alpha}_i = \arg\max \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j k_i k_j \mathbf{x}_i^{\mathbf{T}} \mathbf{x}_j}$$

aby

$$\alpha_i \geq 0$$

$$\sum_{j=1}^{N} \alpha_i k_i = 0$$

Duálny problém

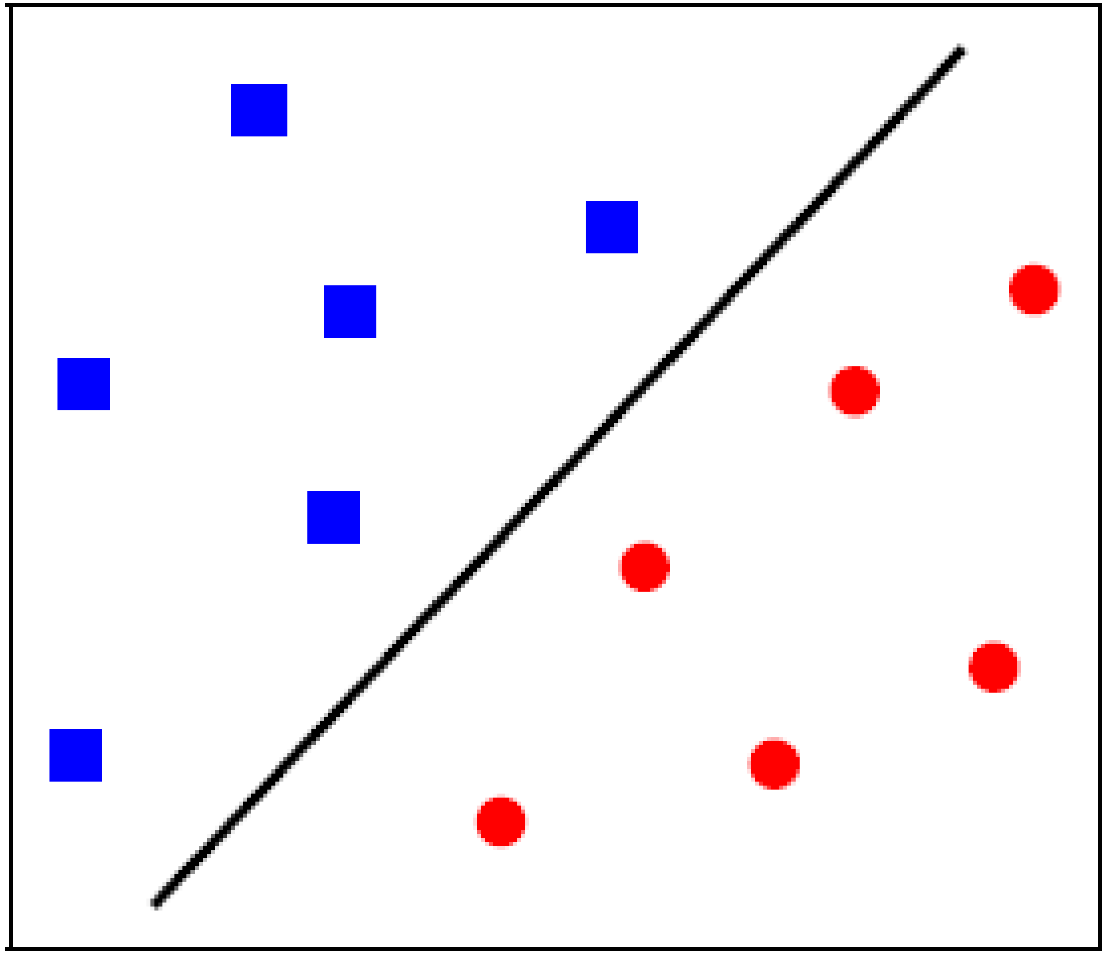
1. Duálnu úlohu vyriešime metódami kvadratic-kého programovania.

2. Určíme množinu podporných vektorov $\boldsymbol{x}_i \in SV$, pre ktoré $\alpha_i > 0$.

3. Určíme hodnotu $b = \frac{1}{|SV|} \sum_{\mathbf{x}_j \in SV} \left(k_j - \sum_{\mathbf{x}_i \in SV} \alpha_i k_i \mathbf{x}_i^T \mathbf{x}_j \right)$

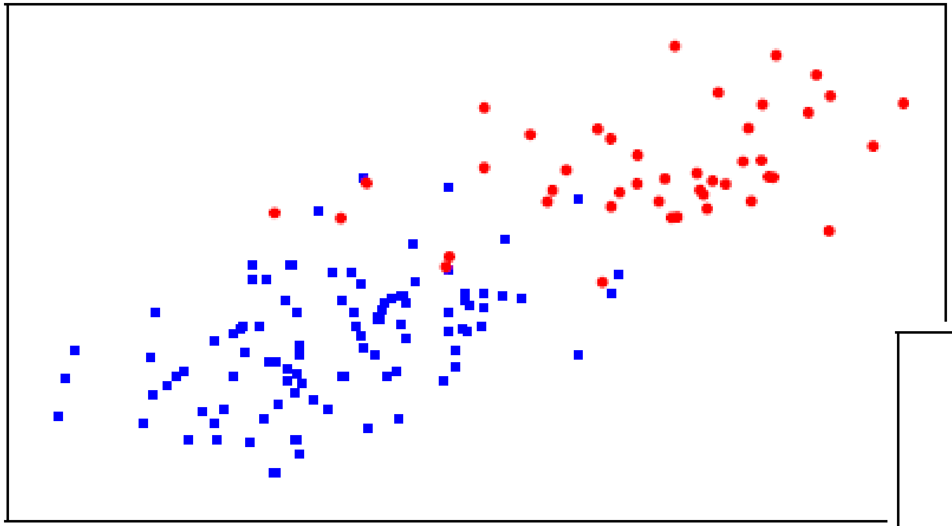
4. Zistíme hodnotu klasifikačnej funkcie $f(\mathbf{x}) = \sum_{\mathbf{x}_i \in SV} \alpha_i k_i \mathbf{x}_i^T \mathbf{x} + b.$

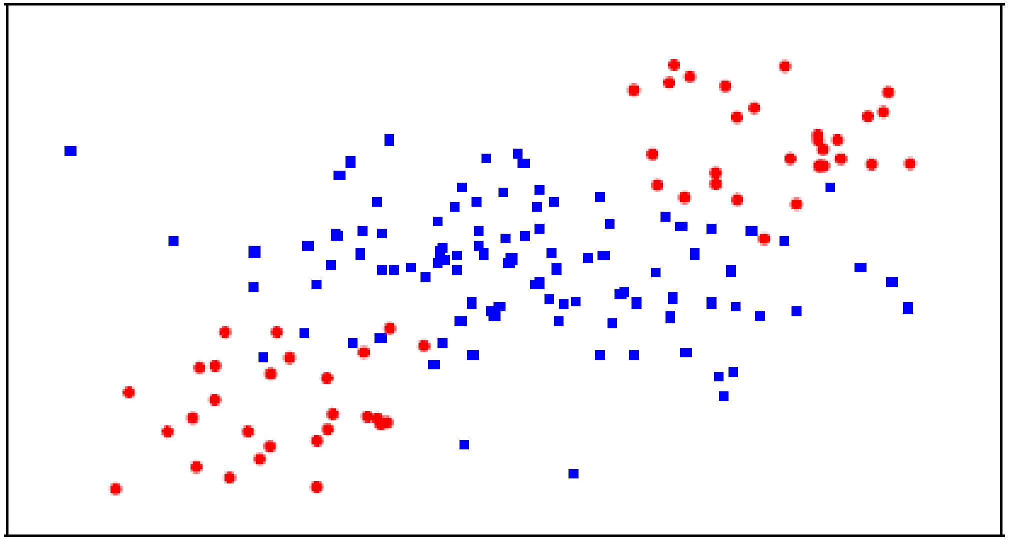
Toto je výsledok



Neseparovateľné dáta

- Dáta separovateľné, ale s malou nenulovou chybou
- Dáta neseparovateľné ani pri nenulovej tolerancii





SVM s voľným rozpätím

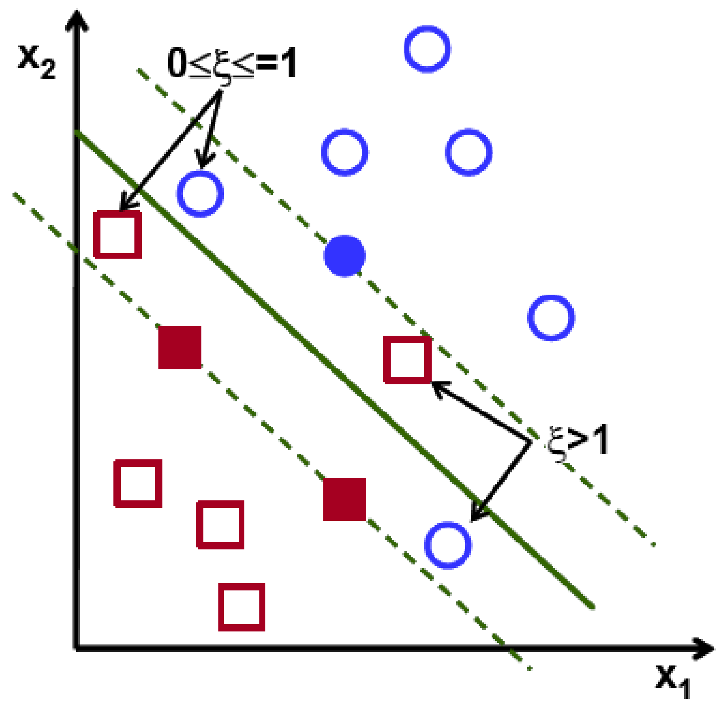
- Pre prvý prípad zavedieme druhotné premenné ξ_i a podmienku nulovej chyby oslabíme

$$(\mathbf{w}^*, b^*, \xi_i^*) = \arg\min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i,$$

aby

$$(\mathbf{w}^T\mathbf{x}_i + b)k_i \geq 1 - \xi_i,$$

- C určuje penalizáciu za chybnú klasifikáciu a ovplyvňuje veľkosť rozpätia
- Lagrangián→duálny problém

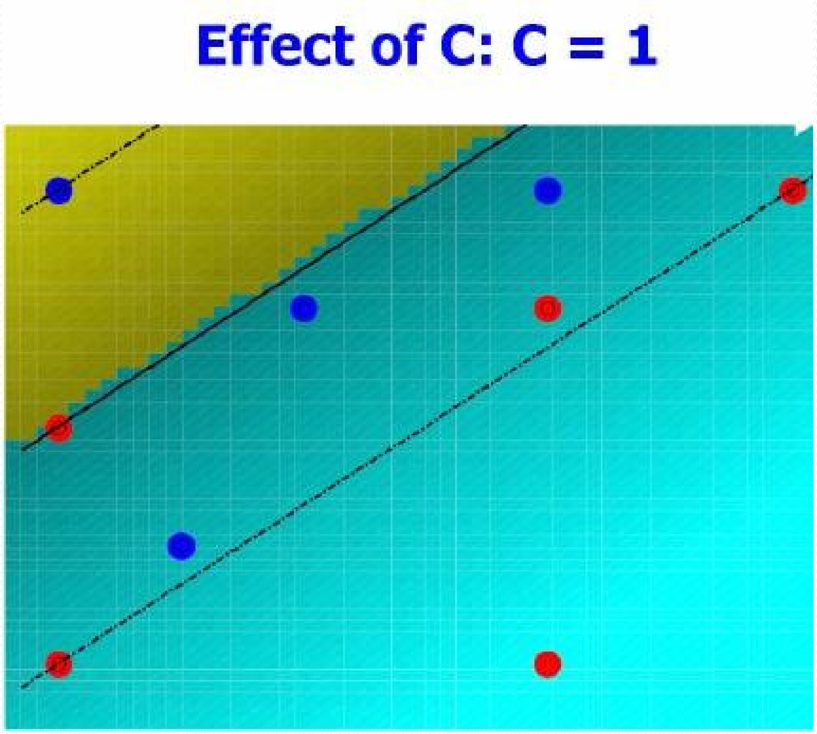


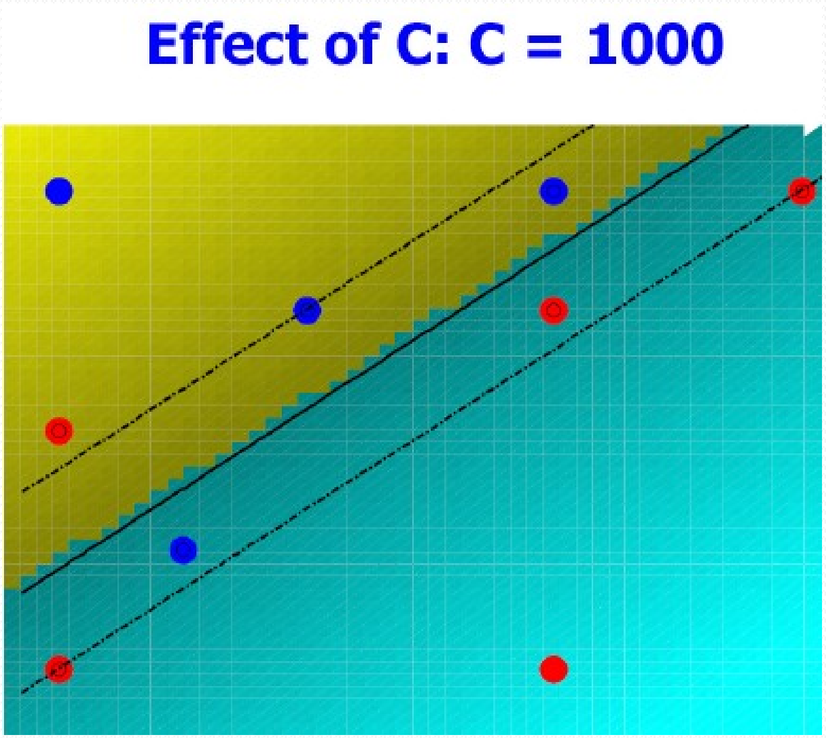
SVM s voľným rozpätím II

- Duálna úloha sa nemení, mení sa podmienka
- Lagrange: $$(\tilde{\mathbf{w}}, \tilde{b}, \tilde{\xi}) = \arg\min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N} \xi_i$$
$$(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b)k_i \geq 1 - \xi_i$$
- Duál: $$\tilde{\alpha}_i = \arg\max \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j k_i k_j \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j$$
$$aby$$
$$0 \leq \alpha_i \leq C$$
$$\sum_{j=1}^{N} \alpha_i k_i = 0$$
$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i k_i \mathbf{x}_i = \sum_{i \in SV} \alpha_i k_i \mathbf{x}_i$$

SVM s voľným rozpätím III

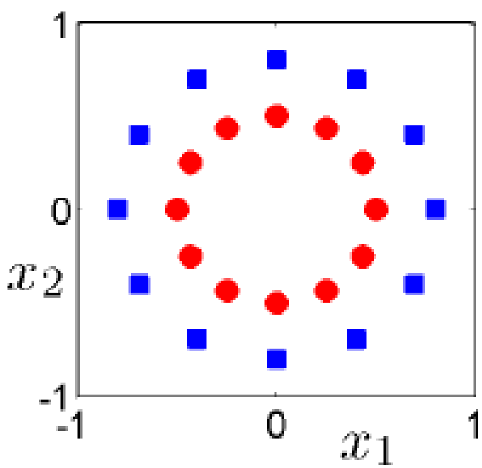
- Vhodnú hodnotu *C* nájdeme vzájomnou validáciou



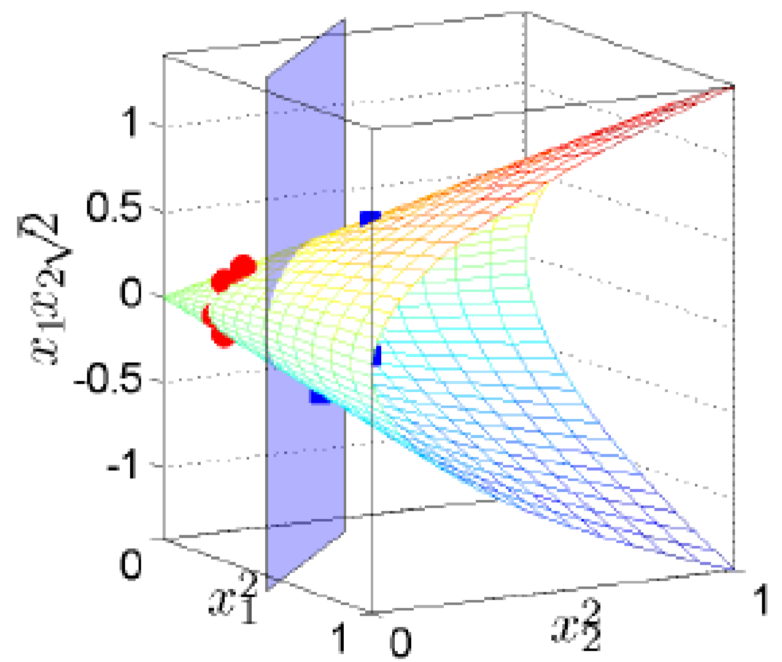


Nelineárny SVM

- Pre lineárne neseparovateľné nájdeme zobra-zenie do iného priestoru, kde to ide lineárne



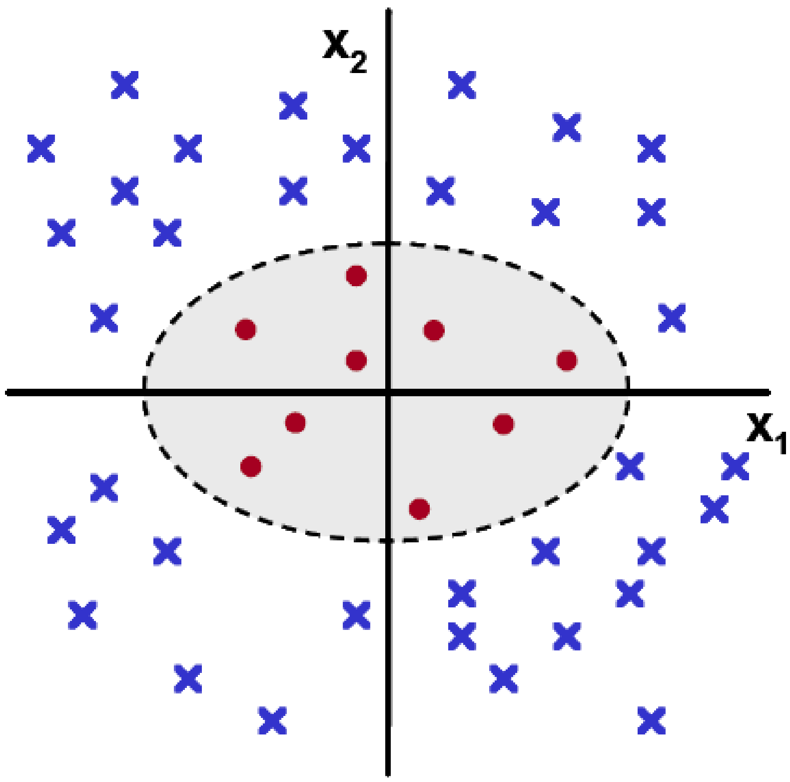
$$\phi : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1 x_2 \end{pmatrix}$$

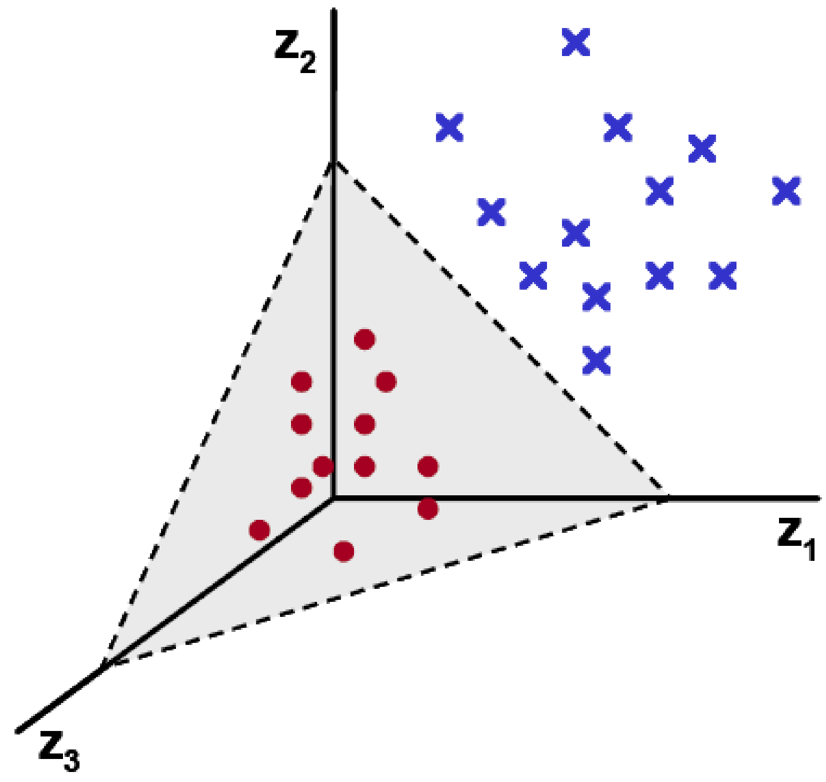


Nelineárny SVM II

$$\varphi : \mathrm{R}^2 \rightarrow \mathrm{R}^3$$

$$\tilde{\alpha}_i = \arg\max \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j k_i k_j \varphi(\mathbf{x}_i)^{\mathrm{T}} \varphi(\mathbf{x}_j)$$





Nelineárny SVM III

1. Duálnu úlohu vyriešime metódami kvadratic-kého programovania.

2. Určíme množinu podporných vektorov $\boldsymbol{x}_i \in SV$, pre ktoré $\alpha_i > 0$.

3. Určíme hodnotu

$$b = \frac{1}{|SV|}\sum_{\mathbf{x}_j \in SV}\left(k_j - \sum_{\mathbf{x}_i \in SV} \alpha_i k_i \boxed{\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)}\right)$$

4. Zistíme hodnotu klasifikačnej funkcie

$$f(\mathbf{x}) = \sum_{\mathbf{x}_i \in SV} \alpha_i k_i \boxed{\phi(\mathbf{x}_i)^T \phi(\mathbf{x})} + b$$

Kernelový trik

- Takže namiesto transformácie každého prízna-kového vektora do viacrozmerného priestoru počítame hodnotu funkcie $K(\ )$ pôvodných vektorov
- Funkcia $K(\boldsymbol{x}, \boldsymbol{y}) = \phi(\boldsymbol{x})^T \phi(\boldsymbol{y})$ sa nazýva kernelová funkcia alebo jadro
- Potom môžeme vzťahy pri výpočte duálnej úlohy zapísať ako:

Kernelový trik II

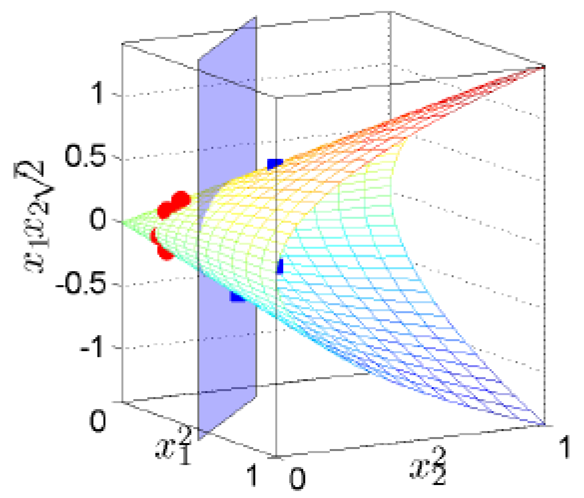
$$\alpha^* = \arg\max\left(\sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j k_i k_j K(\mathbf{x}_i,\mathbf{x}_j)\right)$$

$$b = \frac{1}{|SV|}\sum_{\mathbf{x}_j\in SV}\left(k_j - \sum_{\mathbf{x}_i\in SV}\alpha_i k_i K(\mathbf{x}_i,\mathbf{x}_j)\right)$$

$$f(\mathbf{x}) = \sum_{\mathbf{x}_i\in SV}\alpha_i k_i K(\mathbf{x}_i,\mathbf{x}) + b.$$

Kernelový trik III

- Pre prípad



a zobrazenie

$$\phi : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \end{pmatrix}$$

je kernelová funkcia

$$K(\mathbf{x},\mathbf{y}) = (x_1^2, x_2^2, \sqrt{2}x_1x_2).\begin{pmatrix} y_1^2 \\ y_2^2 \\ \sqrt{2}y_1y_2 \end{pmatrix} = x_1^2y_1^2 + x_2^2y_2^2 + 2x_1y_1x_2y_2 =$$

$$= (x_1y_1 + x_2y_2)^2 =$$

$$= (\mathbf{x}^T\mathbf{y})^2$$

Kernelový trik - príklad

$$K(\mathbf{x}_i,\mathbf{x}_j)=(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)^2=\left(\begin{pmatrix}x_{i1} & x_{i2}\end{pmatrix}\begin{pmatrix}x_{j1}\\ x_{j2}\end{pmatrix}\right)^2=$$

$$=\left(x_{i1}x_{j1}+x_{i2}x_{j2}\right)^2=$$

$$=x_{i1}^2x_{j1}^2+2x_{i1}x_{j1}x_{i2}x_{j2}+x_{i2}^2x_{j2}^2$$

$$\varphi(\mathbf{x})=\begin{pmatrix}x_1^2\\ \sqrt{2}x_1x_2\\ x_2^2\end{pmatrix}\qquad \varphi(\mathbf{x})=\begin{pmatrix}x_1^2\\ x_1x_2\\ x_1x_2\\ x_2^2\end{pmatrix}\qquad \varphi(\mathbf{x})=\frac{1}{\sqrt{2}}\begin{pmatrix}x_1^2-x_2^2\\ 2x_1x_2\\ x_1^2+x_2^2\end{pmatrix}$$

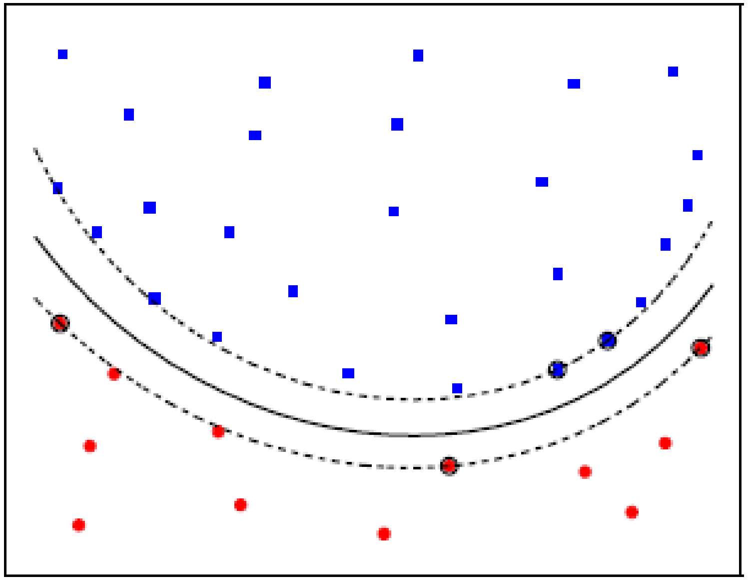
Ako nájdeme jadro

- Ako pre ľubovoľnú funkciu K zistíme, či zodpovedá skalárnemu súčinu $\phi(\boldsymbol{x}_1)^T\phi(\boldsymbol{x}_2)$ v nejakom priestore?
- Konštruujeme špeciálne funkcie – Mercerove jadrá: Ak K je symetrická $(K(\boldsymbol{x}_i, \boldsymbol{x}_j) = K(\boldsymbol{x}_j, \boldsymbol{x}_i))$ pozitívne semidefinitná Gramova matica $\forall \boldsymbol{a} \in \mathbb{R}^N, \boldsymbol{a}K\boldsymbol{a} \geq 0$ a $k_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$, potom K je platné jadro

Príklady jadier

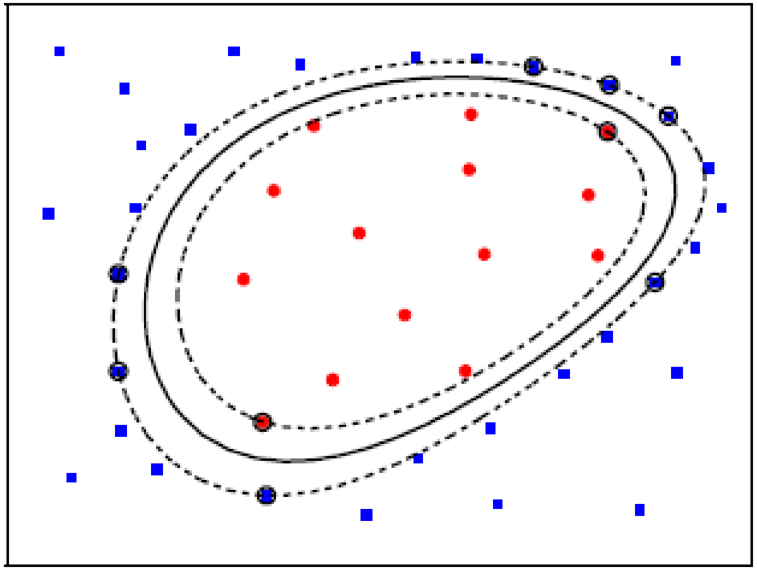
- Polynomiálne jadrá

$$K(\mathbf{x},\mathbf{y}) = (a + \mathbf{x}^T\mathbf{y})^p$$

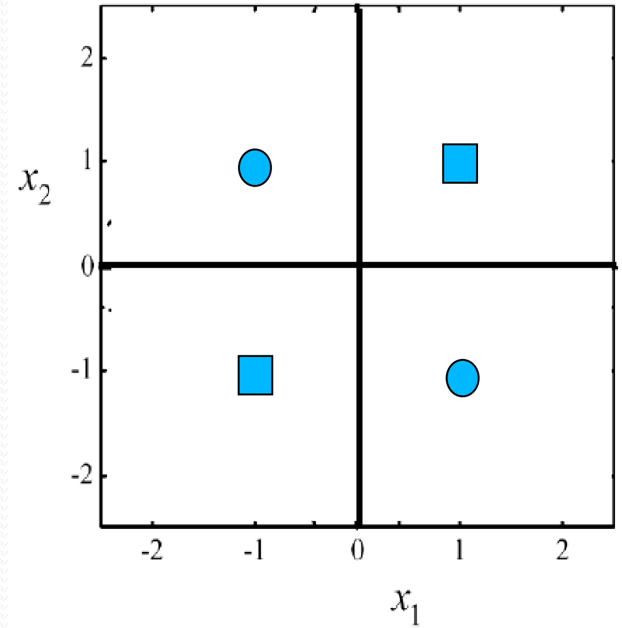


- Gaussovské jadrá

$$K(\mathbf{x},\mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}}$$



Príklad



- Majme triedy

Class 1: $x_1=(-1,-1)$, $x_4=(1,1)$

Class 2: $x_2=(-1,1)$, $x_3=(1,-1)$

$$K(x,x')=(x^T x'+1)^2$$

$$\varphi(x_i)=\begin{bmatrix}1 & \sqrt{2}x_{i,1} & \sqrt{2}x_{i,2} & \sqrt{2}x_{i,1}x_{i,2} & x_{i,1}^2 & x_{i,2}^2\end{bmatrix}^T$$

$$\tilde{\alpha}_i = \arg\max \alpha_1+\alpha_2+\alpha_3+\alpha_4-\frac{1}{2}\sum_{i=1}^{4}\sum_{j=1}^{4}\alpha_i\alpha_j k_i k_j \kappa(\mathbf{x}_i,\mathbf{x}_j)$$

Príklad - pokračovanie

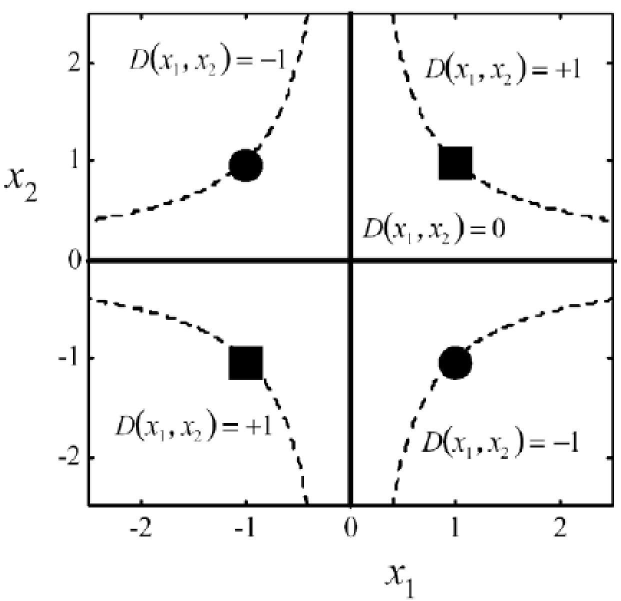
$$9\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 1$$
$$-\alpha_1 + 9\alpha_2 + \alpha_3 - \alpha_4 = 1$$
$$-\alpha_1 + \alpha_2 + 9\alpha_3 - \alpha_4 = 1$$
$$\alpha_1 - \alpha_2 - \alpha_3 + 9\alpha_4 = 1$$

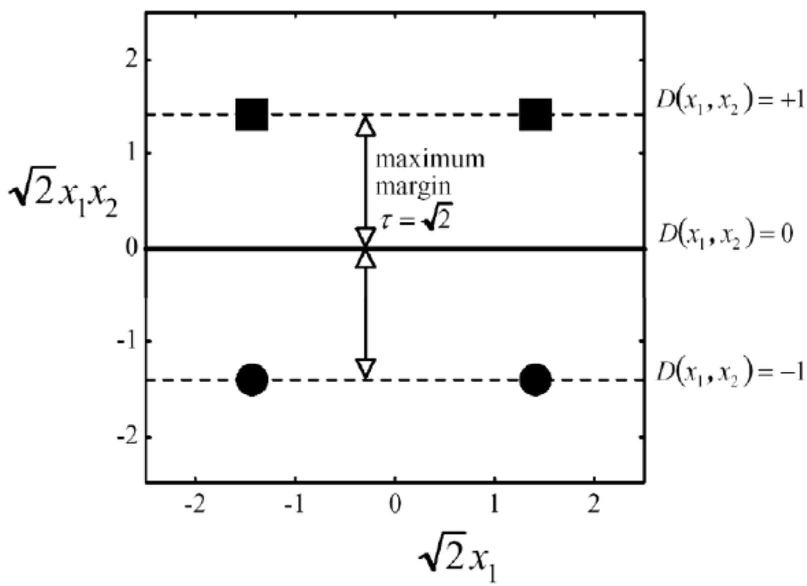
$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.125$$

$$w = \sum_{i=1}^{4} \alpha_i y_i \varphi(x_i) = \begin{bmatrix} 0 & 0 & 0 & 1/\sqrt{2} & 0 & 0 \end{bmatrix}^T$$

$$g(x) = \sum_{i=1}^{6} w_i \varphi_i(x) = x_1 x_2$$

$$y(\mathbf{x}) = \mathbf{w}^{\mathbf{T}} \varphi(\mathbf{x}) + b = \sum_{i=1}^{N} \alpha_i k_i \kappa(\mathbf{x}_i, \mathbf{x}) + b$$





SVM aplikácie

- Kategorizácia textu a hypertextu
- Klasifikácia obrazov
- Bioinformatika (klasifikácia proteínov, detekcia nádorov)
- Rozpoznávanie písaného textu
- ...

Motivačné dáta pre kernel SVM