Lineární klasifikátory

Lineární klasifikátory

- obsah:
  - perceptronový algoritmus
    - základní verze
    - varianta perceptronového algoritmu
    - přihrádkový algoritmus
  - podpůrné vektorové stroje

Lineární klasifikátor

- navrhnout lineární klasifikátor bez ohledu na rozdělení dat
  - výhoda
    - jednoduché
    - typicky výpočetně nenáročné
  - předpoklad
    - trénovací data jsou lineárně klasifikovatelná
  - omezení
    - klasifikace do dvou tříd ω_1 a ω_2 v d-dimenzionálním prostoru

- lineární klasifikátor ... nadrovina

  $g(\mathbf{x}) = \mathbf{w}^T \cdot \mathbf{x} + w_0 = 0$

  kde

  d ... dimenze vstupního prostoru

  $\mathbf{w} = (w_1, \ldots, w_d)$ ... váhový vektor

  w_0 .... práh

Lineární klasifikátor

- navrhnout lineární klasifikátor bez ohledu na rozdělení dat
  - výhoda
    - jednoduché
    - typicky výpočetně nenáročné
  - předpoklad
    - trénovací data jsou lineárně klasifikovatelná
  - omezení
    - klasifikace do dvou tříd ω_1 a ω_2 v d-dimenzionálním prostoru

- lineární klasifikátor ... nadrovina

$$g(\boldsymbol{x}) = \boldsymbol{w}^T \cdot \boldsymbol{x} + w_0 = 0$$

kde

d ... dimenze vstupního prostoru

$\boldsymbol{w} = (w_1, \ldots, w_d)$ ... váhový vektor

w_0 .... práh

Lineární klasifikátor

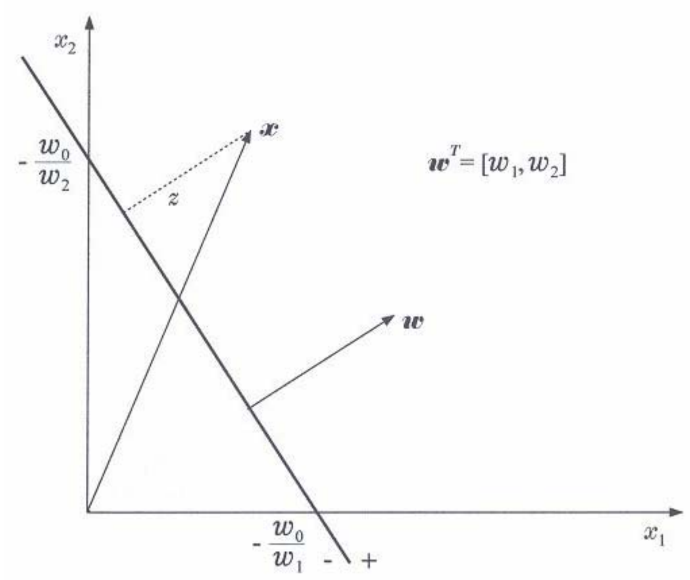
- nadrovina

  $g(\boldsymbol{x}) = \boldsymbol{w}^T \cdot \boldsymbol{x} + w_0 = 0$

- vzdálenost *z* bodu ***x*** od nadroviny

$$z = \frac{|g(x)|}{\|w\|}$$

- rozdělení prostoru na dvě části
  - jedna strana ... $g(\boldsymbol{x}) > 0$
  - druhá strana ... $g(\boldsymbol{x}) < 0$



Perceptronový algoritmus

- cíl
  - nalezení parametrů $\boldsymbol{w}$ a w_0 dělící nadroviny

- lineárně separabilní množiny

  → existuje nadrovina $\boldsymbol{w}^{*T} \cdot \boldsymbol{x} = 0$

  $\boldsymbol{w}^{*T} \cdot \boldsymbol{x} > 0$ pro každý $\boldsymbol{x}$ z ω_1

  $\boldsymbol{w}^{*T} \cdot \boldsymbol{x} < 0$ pro každý $\boldsymbol{x}$ z ω_2

- vztah platí i pro nadroviny neprocházející počátkem

  $\boldsymbol{w}^{*T} \cdot \boldsymbol{x} + w^{*}_{0} = 0$

  - zavede se rozšířený $(d+1)$-dimenzionální prostor

    $\boldsymbol{w'} \equiv (\boldsymbol{w}^{*}, w_0)$

    $\boldsymbol{x'} \equiv (\boldsymbol{x}, 1)$

  → a pak

    $\boldsymbol{w}^{*T} \cdot \boldsymbol{x} + w^{*}_{0} = \boldsymbol{w'}^{T} \cdot \boldsymbol{x'}$

Perceptronový algoritmus

- cíl
  - nalezení parametrů $\boldsymbol{w}$ a w_0 dělící nadroviny

- lineárně separabilní množiny

  → existuje nadrovina $\boldsymbol{w}^{*T}\cdot\boldsymbol{x} = 0$

  $\boldsymbol{w}^{*T}\cdot\boldsymbol{x} > 0$ pro každý $\boldsymbol{x}$ z ω_1

  $\boldsymbol{w}^{*T}\cdot\boldsymbol{x} < 0$ pro každý $\boldsymbol{x}$ z ω_2

- vztah platí i pro nadroviny neprocházející počátkem

  $\boldsymbol{w}^{*T}\cdot\boldsymbol{x} + w^*_0 = 0$

  - zavede se rozšířený $(d+1)$-dimenzionální prostor

    $\boldsymbol{w}' \equiv (\boldsymbol{w}^*, w_0)$

    $\boldsymbol{x}' \equiv (\boldsymbol{x}, 1)$

  → a pak

  $\boldsymbol{w}^{*T}\cdot\boldsymbol{x} + w^*_0 = \boldsymbol{w}'^{T}\cdot\boldsymbol{x}'$

Perceptronový algoritmus – ztrátová funkce

- perceptronový algoritmus
  - příklad optimalizačního problému
    - ztrátová funkce

- ztrátová funkce

$$J(w) = \sum_{x \in Y} \left(\delta_x \cdot w^T \cdot x \right)$$

chybně klasifikované trénovací vzory

  - parametr $\delta_{\boldsymbol{x}}$ → -1 .... pro bod **x** z ω_1
    → $+1$ .... pro bod **x** z ω_2

  - *J(**w**)* nezáporná
    - *J(**w**) = 0* ... *Y* = Ø
    - *J(**w**) > 0* ... (***x*** je z ω_1 & ***x*** klasifikován chybně) OR (***x*** je z ω_2 & ***x*** klasifikován chybně)
  - *J(**w**)* spojitá po částech lineární funkce
    - body, kde se mění počet chybně klasif. vzorů → gradient není definován

Perceptronový algoritmus – odvození 1

- iterativní algoritmus pro minimalizaci ztrátové funkce
  - metoda snižování gradientu

$$w(t+1) = w(t) - \rho_t \frac{\partial J(w)}{\partial w} \Big|_{w=w(t)}$$

w*(t)* ... odhad váhového vektoru v *t*-té iteraci

ρ_t ... posloupnost reálných čísel ... parametr učení

  - derivace ztrátové funkce (tam, kde je definována)

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

  - po dosazení

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

Perceptronový algoritmus – odvození 2

- perceptronový algoritmus

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

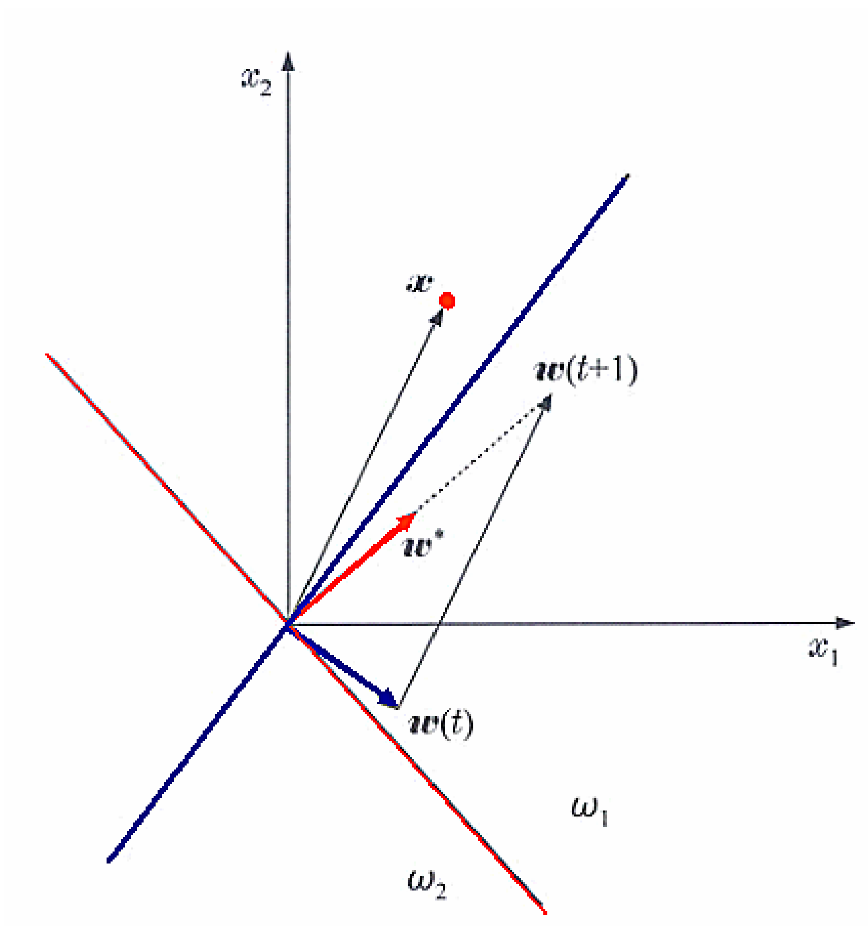
w*(0)* .... libovolně inicializován

$\sum_{x \in Y} \delta_x \cdot x$ ... vektor „opravy“ z nesprávně klasifikovaných vzorů

→ proces učení opakován, dokud nejsou všechny trénovací vzory klasifikovány správně

Perceptronový algoritmus – geometrická interpretace

- geometrická interpretace



Perceptronový algoritmus – konvergence

- podmínky konvergence ... ρ_t

$$\lim_{t\to\infty}\sum_{k=0}^{t}\rho_k = \infty$$

$$\lim_{t\to\infty}\sum_{k=0}^{t}\rho_k^2 < \infty$$

  – např. $\rho_t = c / t$, kde c je konstanta

  – volba ρ_t ... rychlost konvergence → rychlost algoritmu

- perceptronový algoritmus zkonverguje k řešení v konečném počtu kroků
  – důkaz v literatuře

- nalezená řešení
  – nejednoznačná → existuje více než jedna dělicí nadrovina

Perceptronový algoritmus – konvergence

- podmínky konvergence ... ρ_t

$$\lim_{t\to\infty}\sum_{k=0}^{t}\rho_k = \infty$$

$$\lim_{t\to\infty}\sum_{k=0}^{t}\rho_k^2 < \infty$$

  - např. $\rho_t = c\,/\,t$, kde c je konstanta
  - volba ρ_t ... rychlost konvergence → rychlost algoritmu
- perceptronový algoritmus zkonverguje k řešení v konečném počtu kroků
  - důkaz v literatuře
- nalezená řešení
  - nejednoznačná → existuje více než jedna dělicí nadrovina

Perceptronový algoritmus – příklad

- aktuální dělicí přímka (čárkovaná)

  $x_1 + x_2 - 0.5 = 0$

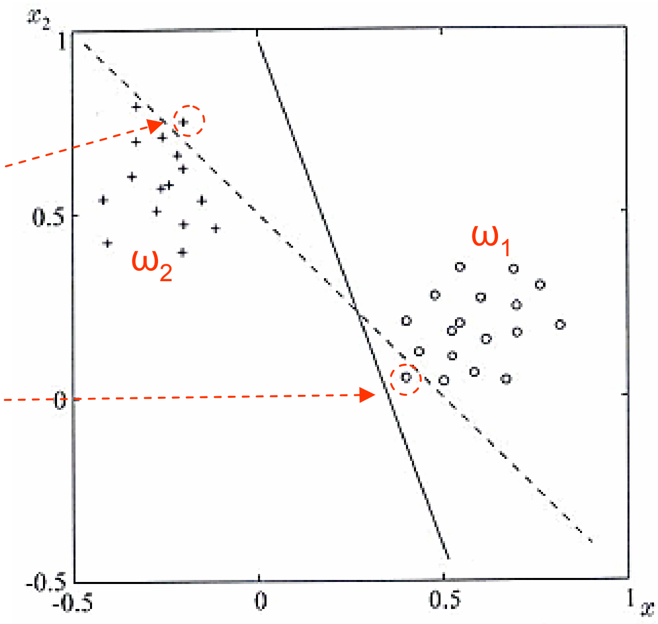
  $\rightarrow$ $\mathbf{w}(t) = (1,1,-0.5)$

- chybně klasifikované vzory

  (-0.2, 0.75)

  (0.4, 0.05)

- aktualizace váhového vektoru ($\rho_t = \rho = 0.7$)



- nová dělicí přímka

  $1.42x_1 + 0.51x_2 - 0.5 = 0$

  $\rightarrow$ správná klasifikace všech trénovacích vzorů

  $\rightarrow$ algoritmus končí

Perceptronový algoritmus – příklad

- aktuální dělicí přímka (čárkovaná)

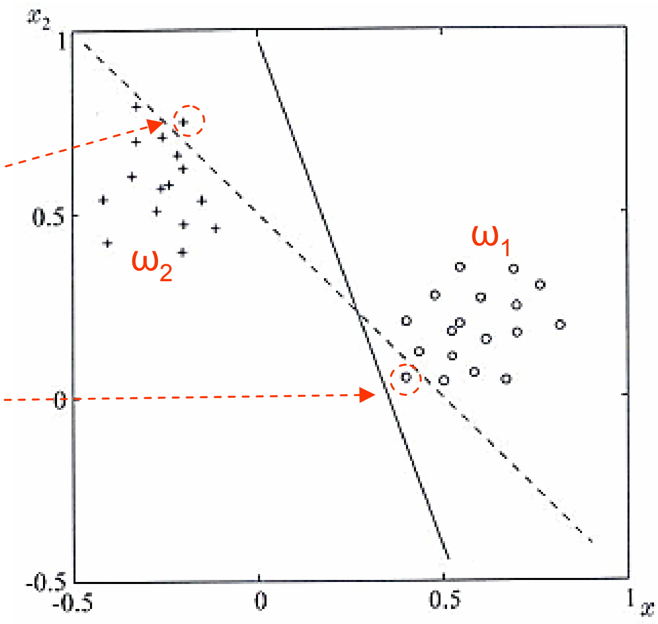
  $x_1 + x_2 - 0.5 = 0$

  $\rightarrow$ **w**(t) = (1,1,-0.5)

- chybně klasifikované vzory

  (-0.2, 0.75)

  (0.4, 0.05)



- aktualizace váhového vektoru ($\rho_t = \rho = 0.7$)

$$\mathbf{w}(t+1) = \begin{pmatrix} 1 \\ 1 \\ -0.5 \end{pmatrix} - 0.7(-1)\begin{pmatrix} 0.4 \\ 0.05 \\ 1 \end{pmatrix} - 0.7(+1)\begin{pmatrix} -0.2 \\ 0.75 \\ 1 \end{pmatrix} = \begin{pmatrix} 1.42 \\ 0.51 \\ -0.5 \end{pmatrix}$$

- nová dělicí přímka

  $1.42x_1 + 0.51x_2 - 0.5 = 0$

  $\rightarrow$ správná klasifikace všech trénovacích vzorů

  $\rightarrow$ algoritmus končí

Variranta perceptronového algoritmu

- modifikace
  - trénovací vzory ... cyklicky předkládány
  - váhový vektor ... změna po každém vzoru
  - opakované předkládání trénovací množiny

- algoritmus
  - aktualizace váhového vektoru

$$w(t+1) = w(t) + \rho \cdot x^t \quad \ldots \quad \text{když } x^t \in \omega_1 \text{ a } w(t) \cdot x^t \leq 0$$

$$w(t+1) = w(t) - \rho \cdot x^t \quad \ldots \quad \text{když } x^t \in \omega_2 \text{ a } w(t) \cdot x^t \geq 0$$

$$w(t+1) = w(t) \quad \ldots \quad \text{jinak}$$

  - konvergence v konečném počtu iterací

vzor $\boldsymbol{x}^t$ klasifikován správně

Perceptron – klasifikace

- klasifikace neznámých vzorů
  - předpoklad: konvergence k váhovému vektoru $\boldsymbol{w}$ a prahu w_0
  - klasifikace

$$\mathbf{w}^T\mathbf{x} + w_0 > 0 \quad \rightarrow \quad \boldsymbol{x} \text{ do } \omega_1$$

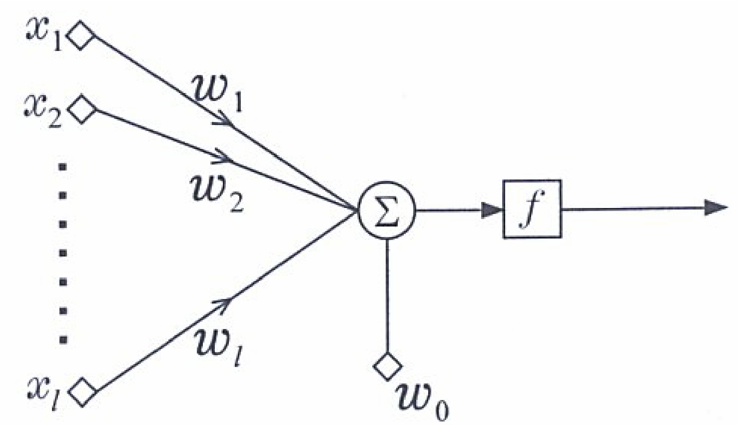
$$\mathbf{w}^T\mathbf{x} + w_0 < 0 \quad \rightarrow \quad \boldsymbol{x} \text{ do } \omega_2$$

→ perceptron

- *f* ... nelineární skoková funkce

$$f(\boldsymbol{\xi}) = +1 \ \ldots \ \boldsymbol{\xi} \geq 0$$

$$f(\boldsymbol{\xi}) = -1 \ \ldots \ \boldsymbol{\xi} < 0$$



Přihrádkový algoritmus

- perceptronový algoritmus učení
  - lineární separabilita tříd
  - předpoklad není splněn → není zaručena konvergence

- přihrádkový algoritmus učení
  - konvergence k optimálnímu řešení i když množiny nejsou lineárně separabilní

Přihrádkový algoritmus

- Krok 1
  - náhodná inicializace $\boldsymbol{w}(0)$
  - přihrádkový vektor $\boldsymbol{w}_S$
  - čítač historie $h_S = 0$
- Krok 2
  - aktualizace $\boldsymbol{w}(t+1)$ podle perceptronového algoritmu
  - *h* .... počet správně klasifikovaných trénovacích vzorů pomocí $\boldsymbol{w}(t+1)$
  - $h > h_S$:

    $\mathbf{w}_S \leftarrow \mathbf{w}(t+1)$

    $h_S \leftarrow h$

- konvergence
  - algoritmus s pravděpodobností 1 konverguje k optimálnímu řešení
    - řešení s nejmenším počtem chybně klasifikovaných trénovacích vzorů

Přihrádkový algoritmus

- Krok 1
  - náhodná inicializace $\boldsymbol{w}(0)$
  - přihrádkový vektor $\boldsymbol{w}_S$
  - čítač historie $h_S = 0$
- Krok 2
  - aktualizace $\boldsymbol{w}(t+1)$ podle perceptronového algoritmu
  - *h ....* počet správně klasifikovaných trénovacích vzorů pomocí $\boldsymbol{w}(t+1)$
  - $h > h_S$

    $\mathbf{w}_S \leftarrow \mathbf{w}(t+1)$

    $h_S \leftarrow h$

- konvergence
  - algoritmus s pravděpodobností 1 konverguje k optimálnímu řešení
    - řešení s nejmenším počtem chybně klasifikovaných trénovacích vzorů

Podpůrné vektorové stroje (SVM)

- support vector machine (SVM)

- vstup
  - lineárně separabilní třídy ω_1 a ω_2
  - trénovací vzory $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$

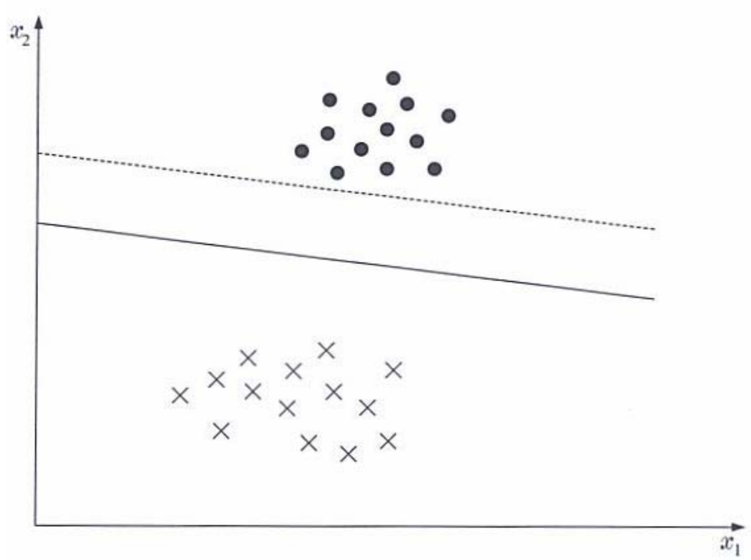
- výstup
  - klasifikační nadrovina $g(\boldsymbol{x})$

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

    → správně klasifikuje trénovací vzory

SVM – idea

- klasifikační nadrovina *g(**x**)* není určena jednoznačně

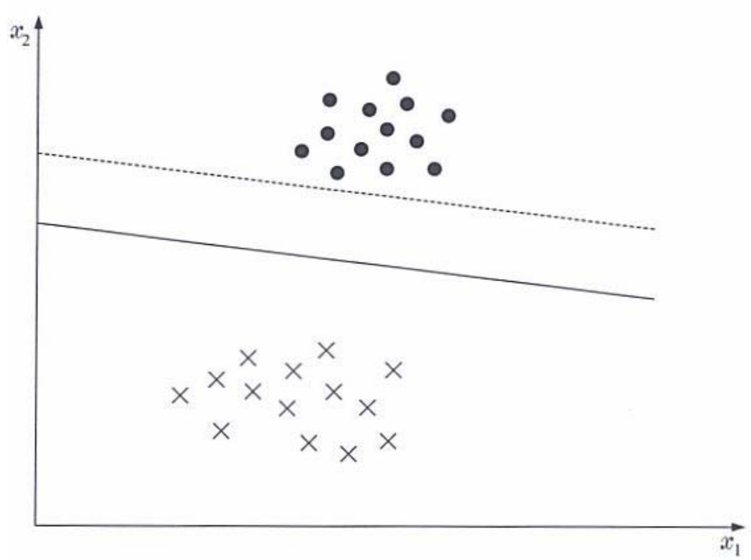


- kterou klasifikační nadrovinu zvolit?
  - „spolehlivá“ práce i s neznámými vzory .... generalizace

→ klasifikace trénovacích dat & „největší okolí“ na obou stranách nadroviny

SVM – idea

- klasifikační nadrovina *g(**x**)* není určena jednoznačně



- kterou klasifikační nadrovinu zvolit?
  - „spolehlivá“ práce i s neznámými vzory .... generalizace

→ klasifikace trénovacích dat & „největší okolí“ na obou stranách nadroviny

SVM – okolí

- umístění klasifikační nadroviny *g(**x**)*
  - žádná třída nemá být zvýhodněná

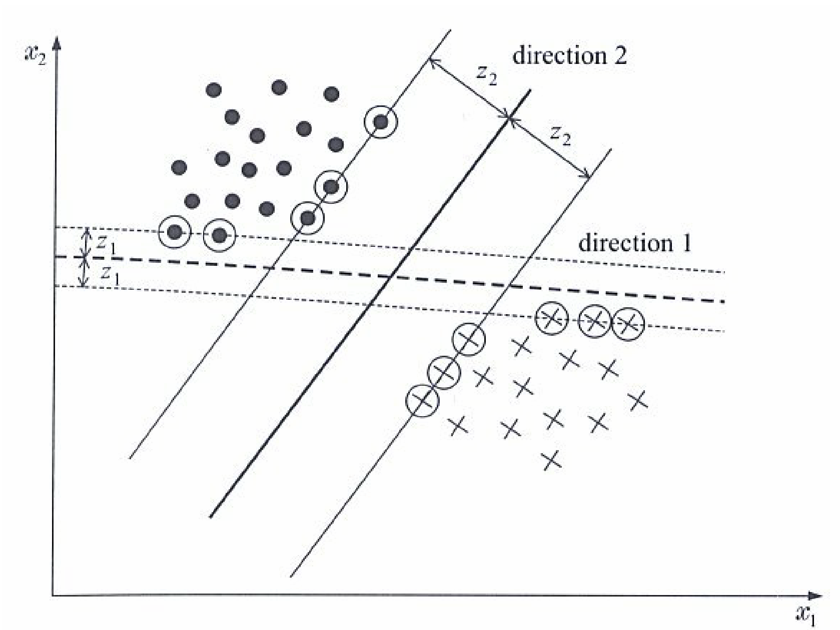
    → stejná vzdálenost (okolí) od nejbližších vzorů v třídách ω_1 a ω_2

→ umístit nadrovinu, aby „okolí“ bylo maximální možné

SVM – okolí

- umístění klasifikační nadroviny *g(**x**)*
  - žádná třída nemá být zvýhodněná

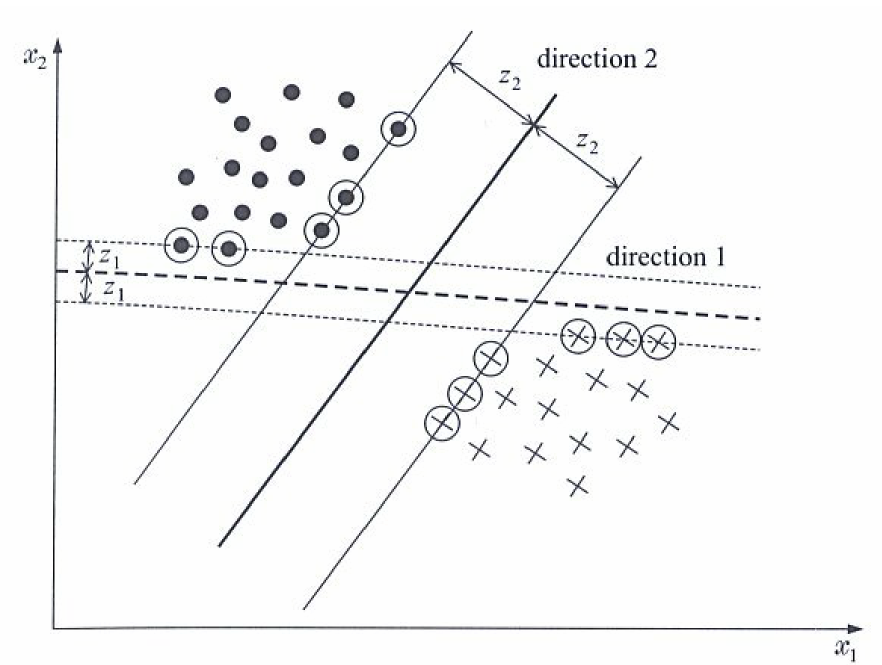
    → stejná vzdálenost (okolí) od nejbližších vzorů v třídách ω_1 a ω_2



→ umístit nadrovinu, aby „okolí“ bylo maximální možné

SVM – okolí

- nadrovina → škálovací koeficient
  - stejný škálovací koeficient → jediná rovnice roviny
  - volba ***w*** a w_0
    - ***x*** nejbližší bod tříd ω_1 → $g(\mathbf{x}) = 1$
    - ***x*** nejbližší bod tříd ω_2 → $g(\mathbf{x}) = -1$



SVM – zápis podmínek

- hledaná klasifikační nadrovina $g(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + w_0 = 0$

1. velikost „okolí“

$$\frac{|g(x_{\omega_1})|}{\|w\|} + \frac{|g(x_{\omega_2})|}{\|w\|} = \frac{2}{\|w\|}$$

2. správná klasifikace trénovacích vzorů

$\mathbf{w}^T\mathbf{x} + w_0 \geq 1$ .... pro trénovací vzor z ω_1

$\mathbf{w}^T\mathbf{x} + w_0 \leq -1$ .... pro trénovací vzor z ω_2

nejbližší bod: $|g(\boldsymbol{x})|=1$

SVM – zápis podmínek

- indikátor třídy vzoru

  y_i → $+1$ .... $\boldsymbol{x_i}$ z ω_1

  → -1 .... $\boldsymbol{x_i}$ z ω_2

- problém
  - parametry $\boldsymbol{w}$ a w_0 nadroviny

    $J(\mathbf{w}) \equiv 1/2 \, || \, \mathbf{w} \, ||^2$ minimální (*)
  - za podmínky

    $y_i(\mathbf{w}^T\mathbf{x} + w_0) \geq 1$ pro $i=1,\ldots,n$ (**)

- poznámky
  - minimalizace normálového vektoru → maximální „okolí“
  - problém kvadratické optimalizace za podmínek lineárních nerovností

SVM – řešení

- problém splňuje Karush-Kuhn-Truckerovy podmínky

  *L(**w**,w_0,**λ**)* ... Lagrangeova funkce

$$L(w, w_0, \lambda) = \frac{1}{2} w^T w - \sum_{i=1}^{n} \lambda_i \left(y_i \left(w^T x + w_0 \right) - 1 \right)$$

→ existuje vektor ***λ*** Lagrangeových multiplikátorů

$$(1) \quad \frac{\partial}{\partial w} L(w, w_0, \lambda) = 0$$

$$(2) \quad \frac{\partial}{\partial w_0} L(w, w_0, \lambda) = 0$$

$$(3) \quad \lambda_i \geq 0 \quad \text{pro } i = 1, \ldots, n$$

$$(4) \quad \lambda_i \left(y_i \left(w^T x + w_0 \right) - 1 \right) = 0 \quad \text{pro } i = 1, \ldots, n$$

- po vyřešení

SVM – řešení

- problém splňuje Karush-Kuhn-Truckerovy podmínky

  *L(**w**,w₀,**λ**)* ... Lagrangeova funkce

$$L(\mathbf{w}, w_0, \boldsymbol{\lambda}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{n} \lambda_i \left(y_i \left(\mathbf{w}^T \mathbf{x} + w_0 \right) - 1 \right)$$

→ existuje vektor ***λ*** Lagrangeových multiplikátorů

$$(1) \quad \frac{\partial}{\partial \mathbf{w}} L(\mathbf{w}, w_0, \boldsymbol{\lambda}) = 0$$

$$(2) \quad \frac{\partial}{\partial w_0} L(\mathbf{w}, w_0, \boldsymbol{\lambda}) = 0$$

$$(3) \quad \lambda_i \geq 0 \quad \text{pro } i = 1, \ldots, n$$

$$(4) \quad \lambda_i \left(y_i \left(\mathbf{w}^T \mathbf{x} + w_0 \right) - 1 \right) = 0 \quad \text{pro } i = 1, \ldots, n$$

- po vyřešení

$$\mathbf{w} = \sum_{i=1}^{n} \lambda_i y_i \mathbf{x}_i \quad \text{a} \quad \sum_{i=1}^{n} \lambda_i y_i = 0$$

SVM – vlastnosti (1a)

- optimální řešení **w**
  - lineární kombinace n_s (n_s<n) trénovacích vzorů, které odpovídají $\lambda_i \neq 0$

$$\mathbf{w} = \sum_{i=1}^{\boxed{n_s}} \lambda_i y_i \mathbf{x}_i$$

    → podpůrné vektory
  - optimální klasifikační nadrovina .... podpůrný vektorový stroj

- podpůrné vektory
  - (4) + $\lambda_i \neq 0$ → podpůrné vektory leží na jedné z nadrovin $\mathbf{w}^T\mathbf{x} + w_0 = \pm 1$
  - podpůrné vektory = trénovací vzory nejbližší klasifikátoru
    → kritické prvky trénovací množiny

SVM – vlastnosti (1b)

- trénovací vzory s $\lambda_i = 0$
  - typicky
    - vně oblasti definované $\boldsymbol{w}^T\boldsymbol{x} + w_0 = \pm 1$
  - degenerované případy
    - na nadrovinách $\boldsymbol{w}^T\boldsymbol{x} + w_0 = \pm 1$

→ výsledná klasifikační nadrovina není citlivá na počet a umístění trénovacích vzorů, které nejsou podpůrnými vektory

SVM – vlastnosti (2)

- vlastnosti problému
  - funkce $J(\mathbf{w})$ striktně konvexní funkce
  - omezující podmínky (**) jsou lineární

→ optimální klasifikační nadrovina jednoznačně určena

SVM – vlastnosti (3)

- dopočtení parametrů z rovnic

$$\mathrm{w} = \sum_{i=1}^{n} \lambda_i y_i x_i \quad \text{a} \quad \sum_{i=1}^{n} \lambda_i y_i = 0$$

  – není snadné → různé algoritmy
    - malé soustavy ... lze dopočítat přímo
    - složitější soustavy ... např. konvexní programování a Lagrangeova dualita

  – typicky více řešení pro λ_i → stejná optimální klasifikační nadrovina

SVM – příklad

- podpůrný vektorový stroj pro body

$$\omega_1 : \begin{bmatrix}1\\1\end{bmatrix} \quad \text{a} \quad \begin{bmatrix}1\\-1\end{bmatrix}$$

$$\omega_2 : \begin{bmatrix}-1\\1\end{bmatrix} \quad \text{a} \quad \begin{bmatrix}-1\\-1\end{bmatrix}$$

– optimální klasifikátor → přímka $g(\mathbf{x}) = x_1 = 0$

- obecně rovnice klasifikační nadroviny

$$g(x) = w_1x_1 + w_2x_2 + w_0 = 0$$

- výsledná optimální nadrovina $g(\mathbf{x}) = x_1 = 0$

$\rightarrow$ $w_0=0$, $w_1=1$, $w_2=0$

→ všechny body podpůrné vektory

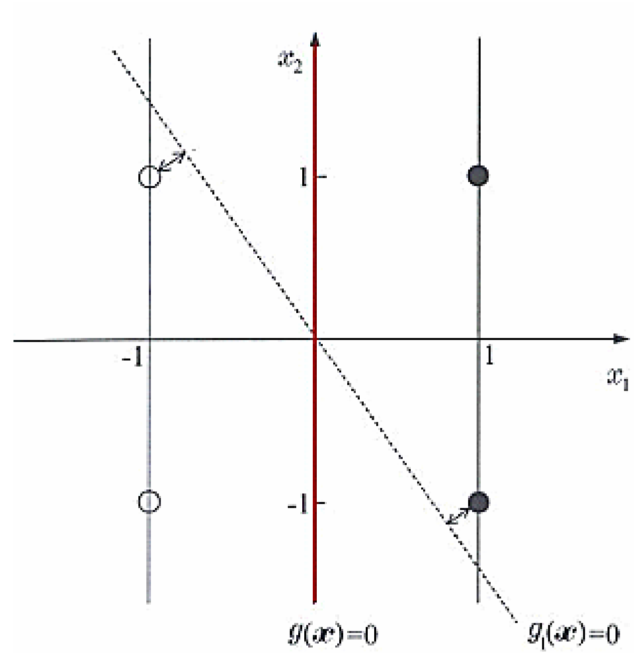
→ velikost „okolí“ 1

SVM – příklad

- podpůrný vektorový stroj pro body

$$\omega_1 : \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ a } \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\omega_2 : \begin{bmatrix} -1 \\ 1 \end{bmatrix} \text{ a } \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$



– optimální klasifikátor → přímka $g(\mathbf{x}) = x_1 = 0$
  - obecně rovnice klasifikační nadroviny

    $$g(\mathbf{x}) = w_1 x_1 + w_2 x_2 + w_0 = 0$$
  - výsledná optimální nadrovina $g(\mathbf{x}) = x_1 = 0$

    → $w_0=0$, $w_1=1$, $w_2=0$

→ všechny body podpůrné vektory

→ velikost „okolí“ 1

SVM – příklad

- analytické řešení w_1, w_2, w_0
  - $J(w) \equiv 1/2\,||w||^2$ minimální
  - za podmínek $y_i(\boldsymbol{w}^T\boldsymbol{x} + w_0) \geq 1$ pro $i=1,\dots,4$

→ dosazení do podmínek

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \dots \quad w_1 + w_2 + w_0 \geq 1$$

$$\mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \dots \quad w_1 - w_2 + w_0 \geq 1$$

$$\mathbf{x}_3 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \dots \quad w_1 - w_2 - w_0 \geq 1$$

$$\mathbf{x}_4 = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \quad \dots \quad w_1 + w_2 - w_0 \geq 1$$

SVM – příklad

- Lagrangeova funkce

$$L(w, w_0, \lambda) = \frac{1}{2} w^T w - \sum_{i=1}^{n} \lambda_i \left(y_i \left(w^T x + w_0 \right) - 1 \right) =$$

$$= \frac{1}{2} \left(w_1^2 + w_2^2 \right) - \lambda_1 \left(w_1 + w_2 + w_0 - 1 \right) - \lambda_2 \left(w_1 - w_2 + w_0 - 1 \right)$$

$$- \lambda_3 \left(w_1 - w_2 - w_0 - 1 \right) - \lambda_4 \left(w_1 + w_2 - w_0 - 1 \right)$$

SVM – příklad

- Karush-Kuhn-Truckerovy podmínky

$$(1)\quad \frac{\partial L}{\partial \mathbf{w}} = 0 \quad \text{je} \quad \frac{\partial L}{\partial w_1} = 0 \quad \ldots \quad \boxed{w_1 = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4}$$

$$\frac{\partial L}{\partial w_2} = 0 \quad \ldots \quad \boxed{w_2 = \lambda_1 - \lambda_2 - \lambda_3 + \lambda_4}$$

$$(2)\quad \frac{\partial L}{\partial w_0} = 0 \quad \ldots \quad \boxed{\lambda_1 + \lambda_2 - \lambda_3 - \lambda_4 = 0}$$

$$(3)\quad \lambda_i \geq 0 \quad \text{pro } i = 1,\ldots,4$$

$$(4)\quad \boxed{\begin{aligned} \lambda_1(w_1 + w_2 + w_0 - 1) &= 0 \\ \lambda_2(w_1 - w_2 + w_0 - 1) &= 0 \\ \lambda_3(w_1 - w_2 - w_0 - 1) &= 0 \\ \lambda_4(w_1 + w_2 - w_0 - 1) &= 0 \end{aligned}}$$

$\rightarrow$ 7 rovnic pro 7 neznámých

SVM – příklad

- řešení 7 rovnic o 7 neznámých není snadné

→ dá více řešení

→ všechny vedou ke stejné optimální klasifikační nadrovině:

$w_1 = 1$

$w_2 = 0$

$w_0 = 0$