Strojno učenje:
8. Stroj potpornih vektora

Jan Šnajder, UNIZG FER, ak. god. 2020./2021.

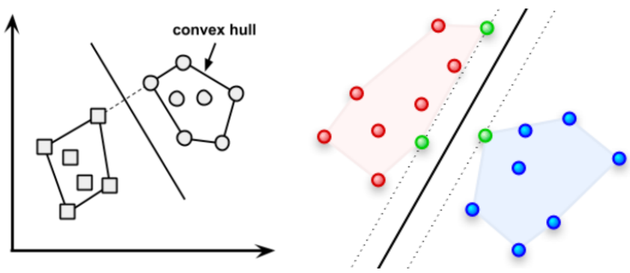
Natuknice za predavanje, v2.2

1 Problem maksimalne margine

- SVM je **linearan model**:

$$h(\mathbf{x}; \mathbf{w}, w_0) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0$$

- Za nelinearnost možemo upotrijebiti preslikavanje $\boldsymbol{\phi}$
- **Margina** – udaljenost između hiperravnine i najbližeg primjera
- SVM nalazi **maksimalnu marginu** $\Rightarrow$ dobra **generalizacija**
- Geometrijski: hiperravnina je simetrala spojice **konveksnih ljusaka** dviju klasa



- Uz pretpostavku **linearne odvojivosti** i uz $y \in \{-1, +1\}$, vrijedi:

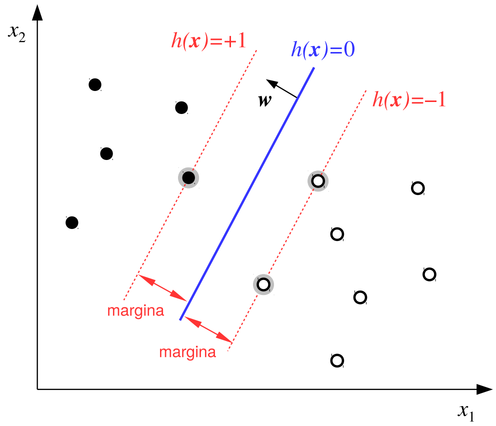
$$\forall (\mathbf{x}^{(i)}, y^{(i)}) \in \mathcal{D}.\ y^{(i)} h(\mathbf{x}^{(i)}) \geqslant 0$$

- Udaljenost primjera $\mathbf{x}^{(i)}$ od hiperravnine je $\frac{1}{\|\mathbf{w}\|} y^{(i)}(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(i)} + w_0)$
- Tražimo hiperravninu maksimalne margine:

$$\underset{\mathbf{w}, w_0}{\operatorname{argmax}} \left\{ \frac{1}{\|\mathbf{w}\|} \min_i \left\{ y^{(i)}(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(i)} + w_0) \right\} \right\}$$

- Vektor $(\mathbf{w}, w_0)$ možemo skalirati tako da za primjere najbliže margini vrijedi:

$$y^{(i)}(\mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0) = 1$$



- Onda za sve primjere vrijedi:

$$y^{(i)}(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(i)} + w_0) \geqslant 1, \qquad i = 1, \dots, N$$

- Optimizacijski problem svodi se na:

$$\underset{\mathbf{w}, w_0}{\operatorname{argmax}} \frac{1}{\|\mathbf{w}\|}$$

uz **ograničenja**:

$$y^{(i)}(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(i)} + w_0) \geqslant 1, \qquad i = 1, \dots, N$$

- Ekvivalentno:

$$\underset{\mathbf{w}, w_0}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{w}\|^2$$

uz ograničenja:

$$y^{(i)}(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(i)} + w_0) \geqslant 1, \quad i = 1, \dots, N$$

$\Rightarrow$ konveksna optimizacija uz ograničenje, preciznije **kvadratno programiranje**

2 Kvadratno programiranje

- Standardni oblik optimizacijskog problema uz ograničenja:

$$\begin{aligned} \text{minimizirati} \quad & f(\mathbf{x}) \\ \text{uz ograničenja} \quad & g_i(\mathbf{x}) \leqslant 0, \quad i = 1, \dots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \dots, p \end{aligned}$$

  - $f : \mathbb{R}^n \to \mathbb{R}$ – ciljna funkcija
  - $h_i : \mathbb{R}^n \to \mathbb{R}$ – ograničenja jednakosti
  - $g_i : \mathbb{R}^n \to \mathbb{R}$ – ograničenja nejednakosti

- Rješivo raznim metodama; mi radimo **Lagrangeovu dualnost + algoritam SMO**

- Omogućava optimizaciju u **dualnoj formi** $\Rightarrow$ SMO, potporni vektori, jezgreni trik

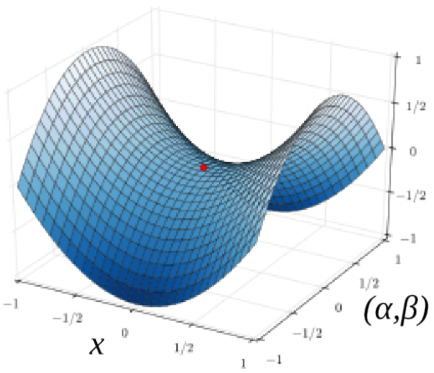
3 Lagrangeova dualnost

- Ograničenja kodiramo u **Lagrangeovu funkciju**:

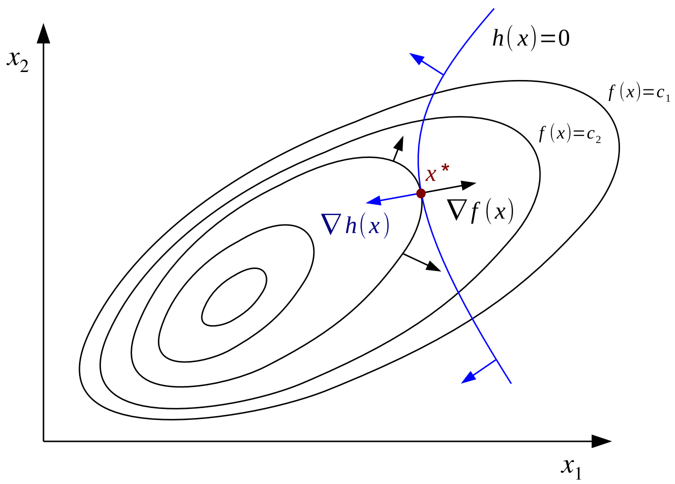
$$L(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = f(\mathbf{x}) + \sum_{i=1}^{m} \alpha_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \beta_i h_i(\mathbf{x})$$

  gdje $\alpha_i \geqslant 0$

- Rješenje originalnog problema je stacionarna točka u kojoj $\nabla L = 0$
- $\nabla L = 0$ je u **točci sedla** funkcije L $\Rightarrow$ minimum po $\mathbf{x}$ a maksimum po $(\boldsymbol{\alpha}, \boldsymbol{\beta})$



- Objašnjenje Lagrangeove funkcije za **ograničenje jednakosti**:



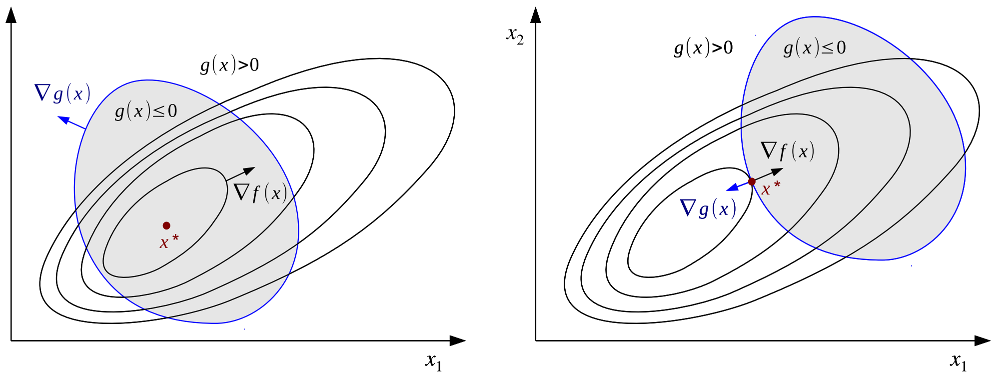
  $\Rightarrow$ u stacionarnoj točci, vektori su kolinearni $\Rightarrow$ postoji β za koju vrijedi:

$$\nabla f(\mathbf{x}) + \beta \nabla h(\mathbf{x}) = 0$$

  što odgovara stacionarnoj točki (Lagrangeove) funkcije:

$$L(\mathbf{x}, \beta) = f(\mathbf{x}) + \beta h(\mathbf{x})$$

- Objašnjenje Lagrangeove funkcije za **ograničenje nejednakosti**:



- Moguća su dva slučaja:
  - minimum je unutar ostvarivog područja $\Rightarrow$ ograničenje nije aktivno ($\alpha = 0$)
  - minimum je izvan ostvarivog područja $\Rightarrow$ za neki $\alpha > 0$ vrijedi:

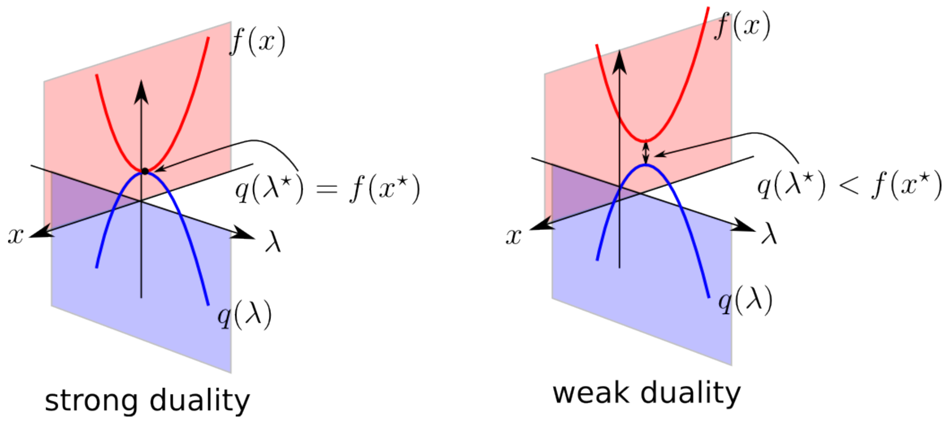
$$\nabla f(\mathbf{x}) = -\alpha \nabla g(\mathbf{x})$$

  što odgovara stacionarnoj točki (Lagrangeove) funkcije:

$$L(\mathbf{x}, \alpha) = f(\mathbf{x}) + \alpha g(\mathbf{x})$$

$\Rightarrow$ u svakom slučaju, za točku rješenja vrijedi $\alpha g(\mathbf{x}) = 0$

- Izvorna ograničenja i dva uvjeta za α čine **Karush-Kuhn-Tuckerove (KKT) uvjete**
- **Načelo dualnosti**: dualni problem je **donja ograda** primarnog problema



- Kod **jake dualnosti** ($f(\mathbf{x})$ konveksna), primarno i dualno rješenje se poklapaju
- **Dualna Lagrangeova funkcija**:

$$\tilde{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\alpha})$$

- Stacionarnu točku od L nalazimo **maksimizacijom** funkcije $\tilde{L}$, tj. dualni problem je:

$$\begin{array}{rl} \text{maksimizirati} & \tilde{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\ \text{uz ograničenja} & \alpha_i \geqslant 0, \quad i = 1, \ldots, m \end{array}$$

4 Optimizacija maksimalne margine

- Lagrangeova funkcija za problem maksimalne margine:

$$L(\mathbf{w}, w_0, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i \Big\{ y^{(i)} \big(\mathbf{w}^{\mathrm{T}} \mathbf{x}^{(i)} + w_0 \big) - 1 \Big\}$$

  gdje $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_N)$, $\alpha_i \geqslant 0$.

- Minimizacija funkcije L po primarnim varijablama:

$$\begin{aligned}
\frac{\partial L(\mathbf{w}, w_0, \boldsymbol{\alpha})}{\partial \mathbf{w}} = 0 \quad &\Rightarrow \quad \mathbf{w} = \sum_{i=1}^{N} \alpha_i y^{(i)} \mathbf{x}^{(i)} \\
\frac{\partial L(\mathbf{w}, w_0, \boldsymbol{\alpha})}{\partial w_0} = 0 \quad &\Rightarrow \quad \sum_{i=1}^{N} \alpha_i y^{(i)} = 0
\end{aligned}$$

- Uvrštavanjem u L dobivamo dualnu Lagrangeovu funkciju:

$$\begin{aligned}
\tilde{L}(\boldsymbol{\alpha}) &= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i \Big\{ y^{(i)} \big(\mathbf{w}^{\mathrm{T}} \mathbf{x}^{(i)} + w_0 \big) - 1 \Big\} \\
&= \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} (\mathbf{x}^{(i)})^{\mathrm{T}} \mathbf{x}^{(j)}
\end{aligned}$$

- Dualni optimizacijski problem SVM-a jest maksimizirati:

$$\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} (\mathbf{x}^{(i)})^{\mathrm{T}} \mathbf{x}^{(j)}$$

  tako da:

$$\begin{aligned}
\alpha_i \geqslant 0, &\quad i = 1, \dots, N \\
\sum_{i=1}^{N} \alpha_i y^{(i)} = 0, &\quad i = 1, \dots, N
\end{aligned}$$

  $\Rightarrow$ kvadratni program rješiv algoritmom **SMO** *(sequential minimal optimization)*

- U točci rješenja vrijede uvjeti KKT:

$$\begin{aligned}
y^{(i)} (\mathbf{w}^{\mathrm{T}} \mathbf{x}^{(i)} + w_0) \geqslant 1, &\quad i = 1, \dots, N \\
\alpha_i \geqslant 0, &\quad i = 1, \dots, N \\
\alpha_i \big(y^{(i)} h(\mathbf{x}^{(i)}) - 1 \big) = 0, &\quad i = 1, \dots, N
\end{aligned}$$

- Od $n + 1$ primarne varijable došli smo na N dualnih varijabli $\Rightarrow$ nekad isplativo

5 Dualni model SVM-a

- Na temelju jednakosti:

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y^{(i)} \mathbf{x}^{(i)}$$

izvodimo **dualnu formulaciju** modela:

$$h(\mathbf{x}) = \underbrace{\mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0}_{\text{Primarno}} = \underbrace{\sum_{i=1}^{N} \alpha_i y^{(i)} \mathbf{x}^{\mathrm{T}} \mathbf{x}^{(i)} + w_0}_{\text{Dualno}}$$

$\Rightarrow$ umjesto težina hiperravnine $\mathbf{w}$, imamo dualne parametre $\boldsymbol{\alpha}$

- Predikcija za $\mathbf{x}^{(i)}$ temelji se na skalarnom umnošku $\mathbf{x}^{\mathrm{T}}\mathbf{x}^{(i)}$ $\Rightarrow$ **sličnost vektora**
- Samo vektori za koje $\alpha_i > 0$ utječu na predikciju $\Rightarrow$ **potporni vektori**
- Težine hiperravnine (primarno) su linearna kombinacija potpornih vektora (dualno):

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y^{(i)} \mathbf{x}^{(i)}$$

- I težina w_0 se može izraziti pomoću potpornih vektora (v. jednadžbu 7.8 u skripti)