9. Stroj potpornih vektora II

Strojno učenje 1, UNIZG FER, ak. god. 2021./2022.

Jan Šnajder, natuknice s predavanja, v2.2

1 Podsjetnik

- Problem maksimalne margine
- Kvadratno programiranje – primarna formulacija
- Lagrangeova dualnost – prijelaz u dualni problem
- Maksimalna margina – dualna formulacija:

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmax}} \Big(\sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y^{(i)} y^{(j)} (\mathbf{x}^{(i)})^\intercal \mathbf{x}^{(j)} \Big)$$

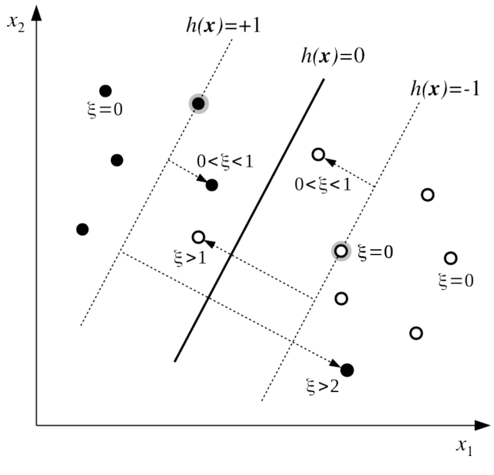
uz ograničenja:

$$\alpha_i \geqslant 0, \quad i = 1, \dots, N$$

$$\sum_{i=1}^N \alpha_i y^{(i)} = 0$$

2 Meka margina

- Gornja formulacija inzistira na **linearnoj odvojivosti** ⇒ uzrokuje **prenaučenost**
- Rješenje: dopustiti ulaske u marginu i pogrešne klasifikacije ⇒ **meka margina**



- Reformulacija ograničenja:

$$y^{(i)}\big(\mathbf{w}^\intercal\mathbf{x}^{(i)} + w_0\big) \geqslant 1 - \xi_i, \quad i = 1, \dots, N$$

gdje $\xi_i \geqslant 0$ govori koliko je primjer $\mathbf{x}^{(i)}$ ušao u marginu, $\xi_i = |y^{(i)} - h(\mathbf{x}^{(i)})|$

  - $\xi_i = 0 \Rightarrow$ ispravno klasificiran i izvan margine
  - $0 < \xi_i \leqslant 1 \Rightarrow$ ispravno klasificiran, ali unutar margine
  - $\xi_i > 1 \Rightarrow$ pogrešno klasificiran

- Ciljnu funkciju proširujemo **kaznom** za primjere za koje $\xi_i > 0$:

$$\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^N \xi_i$$

- $C\uparrow \Rightarrow$ tvrda margina, složen model; $C\downarrow \Rightarrow$ meka margina, jednostavan model
- Optimizacijski problem meke margine:

$$\underset{\mathbf{w},w_0,\boldsymbol{\xi}}{\operatorname{argmin}} \left\{\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^N \xi_i\right\}$$

uz ograničenja:

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

- Pripadna **Lagrangeova funkcija**:

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

- Rješenje zadovoljava **uvjete KKT**:

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

- Minimum po primarnim parametrima:

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

- Uvrštavanjem u L dobivamo **dualnu Lagrangeovu funkciju**:

$$\tilde{L}(\boldsymbol{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} (\mathbf{x}^{(i)})^\intercal \mathbf{x}^{(j)}$$

- Pripadni dualni optimizacijski problem:

$$\underset{\boldsymbol{\alpha}}{\text{argmax}} \Big(\sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} (\mathbf{x}^{(i)})^\intercal \mathbf{x}^{(j)} \Big)$$

  uz ograničenja:

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

- Kao i za tvrdu marginu, uz dodatno ograničenje $\alpha_i \leqslant C$
- Vektori za koje $0 < \alpha_i \leqslant C$ su **potporni vektori** (oni s $\alpha_i = C$ su unutar margine)

3 Gubitak zglobnice

- Alternativna formulacija SVM-a: funkcija gubitka i minimizacija pogreške
- Vrijedi

$$\xi_i = |y^{(i)} - h(\mathbf{x}^{(i)})| = 1 - y^{(i)} h(\mathbf{x}^{(i)})$$

  pa kaznu po primjeru ξ_i možemo napisati kao funkciju gubitka:

$$L(y, h(\mathbf{x})) = \max(0, 1 - y h(\mathbf{x}))\,.$$

  $\Rightarrow$ **gubitak zglobnice (hinge loss)**

- Uvrštavanjem u ciljnu funkciju **primarnog** optimizacijskog problema:

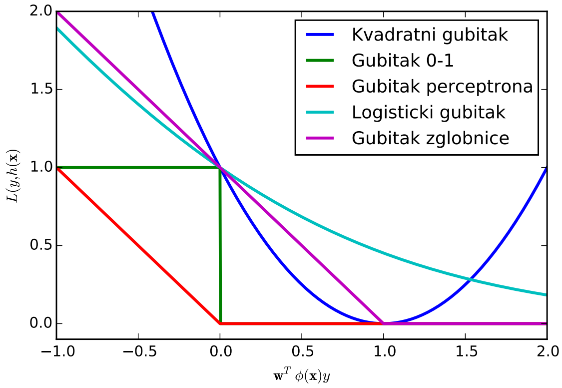
$$\underset{\mathbf{w}, w_0, \boldsymbol{\xi}}{\text{argmin}} \Big\{ \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^{N} \xi_i \Big\}$$

  dobivamo:

$$E(\mathbf{w}|\mathcal{D}) = \sum_{i=1}^{N} \max\big(0, 1 - y^{(i)} h(\mathbf{x}^{(i)})\big) + \frac{\lambda}{2}\|\mathbf{w}\|^2$$

  gdje $\lambda = 1/C$

- Može se minimizirati npr. stohastičkim (pod)gradijentnim spustom
- Usporedba funkcija gubitaka:



4 Alternativni SVM algoritmi

- Primarna formulacija $\Rightarrow$
  - Gubitak zglobnice $\Rightarrow$ SGD, SGP, koordinatni spust, ...
  - Kvadratno programiranje (QP) $\Rightarrow$
    * Lagrangeova dualnost $\Rightarrow$ Dualno QP $\Rightarrow$ SMO
    * Koordinatni spust, metode kazne, metode unutarnje točke,...
- SVM rješavači: SMO, LibSVM, LibLinear, SVM$^{\text{light}}$, Pegasos,...

5 Napomene

- SVM regresija (SVR)
- Hiperparametar C – određuje složenost, odabrati unakrsnom provjerom
- Skaliranje – skalirati značajke, da ne dominiraju one s većim rasponom
- Višeklasna klasifikacija – preporuča se OVO zbog manje neuravnoteženosti klasa
- Probabilistički izlaz – može se aproksimirati Plattovom metodom (izlazna sigmoida)

$$P(y = 1|\mathbf{x}) = \sigma(ah(\mathbf{x}) + b)$$

- Nelinearnost – preslikavanjem $\boldsymbol{\phi}$ ili **jezgrenim trikom** $\Rightarrow$ iduće predavanje