10. Jezgrene metode

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

Jan Šnajder, natuknice s predavanja, v1.3

1 Jezgrene funkcije

- Umjesto težina uz vektor značajki $\mathbf{x}$, izračunavamo **sličnost** dvaju primjera
- Naročito prikladno kada se primjeri teško vektoriziraju (npr. jer imaju strukturu)
- **Jezgrena funkcija** (*kernel function*): $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
- Jezgrena funkcija je **mjera sličnosti** ako zadovoljava:
  - $\kappa(\mathbf{x}, \mathbf{x}) = 1$
  - $0 \leqslant \kappa(\mathbf{x}, \mathbf{x}') \leqslant 1$
  - $\kappa(\mathbf{x}, \mathbf{x}') = \kappa(\mathbf{x}', \mathbf{x})$
- Jezgre za strukturirane podatke: **string kernels**, **tree kernels**, **graph kernels**
- Tipične jezgrene funkcije za primjere u vektorskom prostoru:
  - **Linearna jezgra**: $\kappa(\mathbf{x}, \mathbf{x}') = \mathbf{x}^{\mathrm{T}}\mathbf{x}'$
  - **Radijalna bazna funkcija (RBF)**: općenito jezgra tipa $\kappa(\mathbf{x}, \mathbf{x}') = f(\|\mathbf{x} - \mathbf{x}'\|)$
  - **Gaussova RBF-jezgra**:

$$\kappa(\mathbf{x}, \mathbf{x}') = \exp\Big(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\Big) = \exp\big(-\gamma\|\mathbf{x} - \mathbf{x}'\|^2\big)$$

    gdje je σ^2 je širina pojasa (*bandwidth*), $\gamma = 1/2\sigma^2$ je preciznost (manja σ^2 $\Leftrightarrow$ veća γ $\Leftrightarrow$ primjeri su međusobno sve različitiji)
  - **Ekponencijalna jezgra**: $\kappa(\mathbf{x}, \mathbf{x}') = \exp\big(-\gamma\|\mathbf{x} - \mathbf{x}'\|\big)$
  - **Inverzna kvadratna jezgra**: $\kappa(\mathbf{x}, \mathbf{x}') = \frac{1}{1+\|\mathbf{x}-\mathbf{x}'\|^2}$
- Umjesto euklidske, može se koristiti **Mahalanobisova udaljenost**:

$$\kappa(\mathbf{x}, \mathbf{x}') = \exp\Big(-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{x}')\Big)$$

  gdje je $\mathbf{\Sigma}$ kovarijacijska matrica značajki

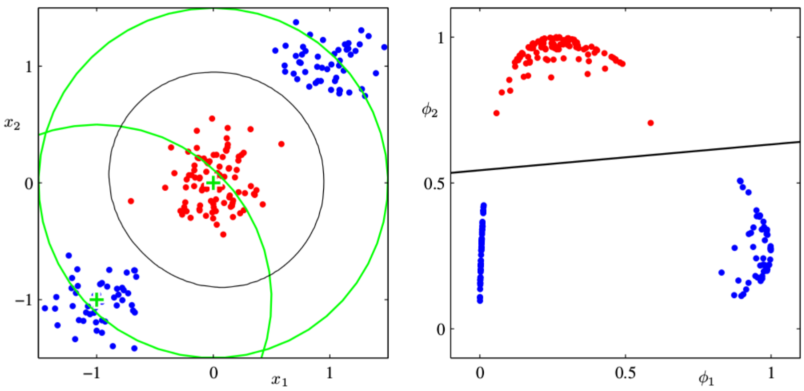
2 Jezgreni strojevi

- Preslikavanje $\boldsymbol{\phi}$ koje za bazne funkcije ϕ_j koristi jezgrene funkcije:

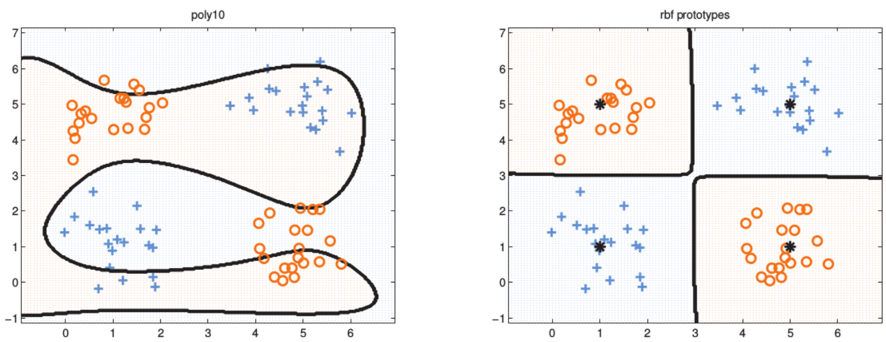
$$\boldsymbol{\phi}(\mathbf{x}) = \big(1, \kappa(\mathbf{x}, \boldsymbol{\mu}_1), \kappa(\mathbf{x}, \boldsymbol{\mu}_2), \dots, \kappa(\mathbf{x}, \boldsymbol{\mu}_m)\big)$$

  gdje su $\boldsymbol{\mu}_j \in \mathcal{X}$ odabrane točke u prostoru primjera

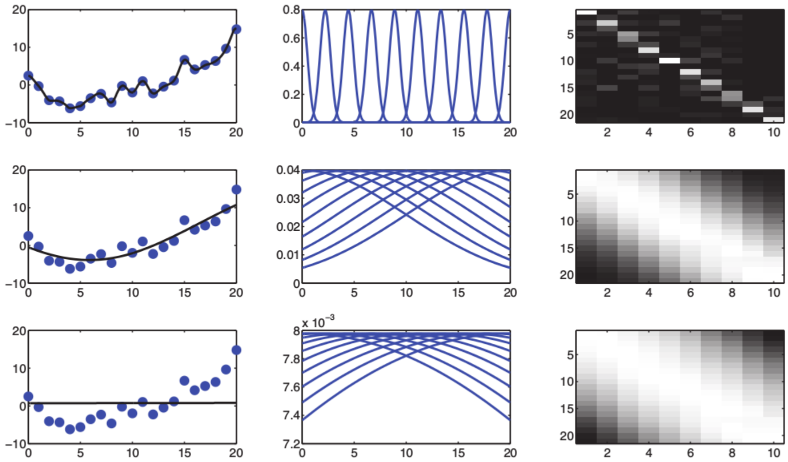
- **Jezgreni stroj** (*kernel machine*) – poopćeni linearni model s takvim preslikavanjem
- Primjer (iz MLPP) – klasifikacija:



- Primjer (iz MLPP) – klasifikacija:



- Primjer (iz MLPP) – regresija:



- Uniforman raspored $\boldsymbol{\mu}_j \Rightarrow$ neprilagođen podatcima, problem visokih dimenzija
- Alternativa: $\boldsymbol{\mu}_j$ su primjeri iz skupa za učenje:

$$\boldsymbol{\phi}(\mathbf{x}) = (1, \kappa(\mathbf{x}, \mathbf{x}_1), \kappa(\mathbf{x}, \mathbf{x}_2), \dots, \kappa(\mathbf{x}, \mathbf{x}_N))$$

- Problem: primjera može biti puno; rješenje: **L1-regularizacija**
- **Rijetki jezgreni strojevi**: L1VM, SVM

3 Jezgreni trik

- SVM je rijedak jezgreni stroj, no umjesto preslikavanja koristi jezgreni trik
- **Jezgreni trik** – skalarni produkt vektora zamjenjuje se jezgrenom funkcijom:

$$\kappa(\mathbf{x}, \mathbf{x}') = \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}')$$

- Model SVM:

$$h(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i y^{(i)} \boldsymbol{\phi}(\mathbf{x})^{\mathsf{T}} \boldsymbol{\phi}(\mathbf{x}^{(i)}) + w_0 = \sum_{i=1}^{N} \alpha_i y^{(i)} \kappa(\mathbf{x}, \mathbf{x}^{(i)}) + w_0$$

- Ciljna funkcija (kvadratno programiranje):

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

- **Inverzno oblikovanje** – odabiremo jezgru i time implicitno definiramo $\boldsymbol{\phi}$
- Prednosti:
  - Manja računalna složenost (izračun κ je često jeftiniji od izračuna $\boldsymbol{\phi}$)
  - Nekad je lakše definirati κ nego $\boldsymbol{\phi}$ (strukturirani podatci: nizovi, stabla, grafovi)

- Prostor koji inducira κ može biti visoko (potencijalno beskonačno) dimenzijski

- Uvjet: κ odgovara skalarnom produktu u nekom vekt. prostoru $\Rightarrow$ **Mercerova jezgra**
- Jezgrena matrica (*kernel matrix*):

$$\mathbf{K} = \begin{pmatrix} \kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) & \kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) & \dots & \kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(N)}) \\ \kappa(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) & \kappa(\mathbf{x}^{(2)}, \mathbf{x}^{(2)}) & \dots & \kappa(\mathbf{x}^{(2)}, \mathbf{x}^{(N)}) \\ \vdots & \vdots & \ddots & \vdots \\ \kappa(\mathbf{x}^{(N)}, \mathbf{x}^{(1)}) & \kappa(\mathbf{x}^{(N)}, \mathbf{x}^{(2)}) & \dots & \kappa(\mathbf{x}^{(N)}, \mathbf{x}^{(N)}) \end{pmatrix}$$

- $\mathbf{K} = \mathbf{\Phi}\mathbf{\Phi}^{\mathrm{T}} \Leftrightarrow \mathbf{K}$ je **Gramova matrica** (matrica skalarnih produkata)
- Gramova matrica je uvijek pozitivno semidefinitna ($\forall \mathbf{x}.\ \mathbf{x}^{\mathrm{T}}\mathbf{K}\mathbf{x} \geqslant 0$)
- **Mercerov teorem**: $\mathbf{K}$ je pozitivno semidefinitna $\Leftrightarrow \kappa$ je Mercerova jezgra
- Inducirani prostor: skalarni produkt + proizvoljna dimenzija $\Rightarrow$ **Hilbertov prostor**
- Mercerove jezgre: linearna, polinomijalna, RBF-jezgra, string kernels, ...
- Preslikavanje polinomijalne jezgre
  - Općenito: $\kappa(\mathbf{x}, \mathbf{x}') = (\gamma \mathbf{x}^{\mathrm{T}}\mathbf{x}' + c)^d$
  - Primjer za $n = 2$, $d = 2$, $c = 0$, $\gamma = 1$:

$$\begin{aligned} \kappa(\mathbf{x}, \mathbf{z}) &= (\mathbf{x}^{\mathrm{T}}\mathbf{z})^2 = ((x_1, x_2)^{\mathrm{T}}(z_1, z_2))^2 = (x_1 z_1 + x_2 z_2)^2 \\ &= (x_1 z_1)^2 + 2(x_1 z_1)(x_2 z_2) + (x_2 z_2)^2 = x_1^2 z_1^2 + \sqrt{2} x_1 x_2 \sqrt{2} z_1 z_2 + x_2^2 z_2^2 \\ &= (x_1^2, \sqrt{2} x_1 x_2, x_2^2)^{\mathrm{T}}(z_1^2, \sqrt{2} z_1 z_2, z_2^2) = \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{z}) \end{aligned}$$

$$\Rightarrow \boldsymbol{\phi}(\mathbf{x}) = (x_1^2, \sqrt{2} x_1 x_2, x_2^2)$$

- Preslikavanje RBF-jezgre:
  - $\mathbf{K}$ je punog ranga $\Leftrightarrow \boldsymbol{\phi}(\mathbf{x})$ su lin. nezavisni $\Rightarrow$ **beskonačnodimenzijski prostor**
  - $\gamma \to \infty \Leftrightarrow \kappa(\mathbf{x}, \mathbf{x}') \to 0 \Leftrightarrow \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}) = 0 \Leftrightarrow$ primjeri su ortonormirani $\Leftrightarrow$ primjeri su vrhovi višedimenzijskog simpleksa $\Leftrightarrow$ linearno su odvojivi
- Složenije Mercerove jezgre gradimo operacijama koje zadržavaju to svojstvo:

$$\begin{aligned} \kappa(\mathbf{x}, \mathbf{x}') &= \alpha \kappa_1(\mathbf{x}, \mathbf{x}') && \alpha > 0 \\ \kappa(\mathbf{x}, \mathbf{x}') &= f(\mathbf{x})\kappa_1(\mathbf{x}, \mathbf{x}')f(\mathbf{x}') && f \text{ – bilo koja funkcija} \\ \kappa(\mathbf{x}, \mathbf{x}') &= q\big(\kappa_1(\mathbf{x}, \mathbf{x}')\big) && q \text{ – polinom s poz. koef.} \\ \kappa(\mathbf{x}, \mathbf{x}') &= \exp\big(\kappa_1(\mathbf{x}, \mathbf{x}')\big) && \\ \kappa(\mathbf{x}, \mathbf{x}') &= \kappa_1\big(\boldsymbol{\phi}(\mathbf{x}), \boldsymbol{\phi}(\mathbf{x})\big) && \boldsymbol{\phi} : \mathbb{R}^n \to \mathbb{R}^{m+1} \\ \kappa(\mathbf{x}, \mathbf{x}') &= \kappa_1(\mathbf{x}, \mathbf{x}') + \kappa_2(\mathbf{x}, \mathbf{x}') && \\ \kappa(\mathbf{x}, \mathbf{x}') &= \kappa_1(\mathbf{x}, \mathbf{x}')\kappa_2(\mathbf{x}, \mathbf{x}') && \end{aligned}$$

$\Rightarrow$ **multiple kernel learning (MKL)**

4 Napomene

- Odabir modela kod SVM-a
  - RBF-jezgra: hiperparametri C i γ su međuovisni ($\gamma\uparrow \Leftrightarrow C\downarrow$)
  - odabir modela najčešće se radi **pretraživanjem po rešetci** (*grid search*)
- Linearna jezgra – ne daje nelinearnost, ali daje rijetka rješenja (potporni vektori)
- Jezgeni trik primjenjiv je na druge algoritme (npr. kernelizirana linearna regresija)
- **Aproksimacija kernela** (kada je N velik) – aproksimacija preslikavanja $\boldsymbol{\phi}$ + SGD