10. Jezgrene metode

Strojno učenje 1, UNIZG FER, ak. god. 2023./2024.

Jan Šnajder, vježbe, v1.11

1 Zadatci za učenje

1. [***Svrha: Znati definirati osnovne jezgrene funkcije. Znati definirati jezgreni stroj i razumjeti razliku između jezgrenog stroja i rijetkog jezgrenog stroja.***]

   (a) Definirajte jezgrenu funkciju, RBF-jezgru i Gaussovu jezgru.

   (b) Je su li RBF-jezgre osjetljive na razlike u skalama značajki? Zašto?

   (c) Definirajte Mahalanobisovu udaljenost i RBF-jezgru koja koristi tu udaljenost. Navedite primjer u kojemu biste koristili tu jezgru umjesto Gaussove jezgre.

   (d) Definirajte jezgreni stroj i rijetki jezgreni (vektorski) stroj. Koji od njih je parametarski a koji neparametarski algoritam i što to znači?

2. [***Svrha: Isprobati preslikavanje primjera u prostor značajki primjenom Gaussovih baznih funkcija. Razumjeti kako preslikavanje utječe na broj parametara i hiperparamera modela.***] Raspolažemo skupom primjera:

$$\mathcal{D} = \big\{(\mathbf{x}^{(i)}, y^{(i)})\big\}_{i=1}^{5} = \big\{\big((-1,-1),0\big), \big((0,0),0\big), \big((3,-3),1\big), \big((-2,1),1\big), \big((-4,2),1\big)\big\}\,.$$

   (a) U ulaznome prostoru skicirajte diskriminacijsku granicu $h(\mathbf{x}) = 0$ koju biste dobili logističkom regresijom uz $\boldsymbol{\phi}(\mathbf{x}) = (1, \mathbf{x})$, tj. bez preslikavanja (izračun nije potreban).

   (b) Na isti skup primjera primijenite jezgreni stroj s baznim funkcijama:

$$\phi_j(\mathbf{x}) = \exp\Big(-\frac{\|\mathbf{x}-\boldsymbol{\mu}_j\|^2}{2\sigma_j^2}\Big)\,.$$

   Konkretno, koristite dvije bazne funkcije s parametrima $\boldsymbol{\mu}_1 = (0,0)$, $\boldsymbol{\mu}_2 = (-3,3)$ i $\sigma_1 = \sigma_2 = 1$. Skicirajte primjere u prostoru značajki (dimenzije ϕ_1 i ϕ_2) i granicu koju biste dobili logističkom regresijom (izračun nije potreban).

   (c) Koliko ovaj jezgreni stroj ima parametara a koliko hiperparametara? Kako biste u praksi odredili vrijednosti hiperparametara modela? Određuju li u ovom slučaju hiperparametri složenost modela? Obrazložite odgovor.

3. [***Svrha: Razumjeti jezgreni trik kod SVM-a.***]

   (a) Za klasifikaciju primjera u ulaznome prostoru $X = \mathbb{R}^2$ koristimo polinomijalnu jezgrenu funkciju $\kappa(\mathbf{x},\mathbf{z}) = (\mathbf{x}^\mathrm{T}\mathbf{z}+1)^2$. Pokažite da je za $n = 2$ jezgra κ Mercerova jezgra. Zašto je to bitno?

   (b) Izvedite pripadno preslikavanje $\boldsymbol{\phi}(\mathbf{x})$ za $n = 2$. U koji će vektor u prostoru značajki efektivno biti preslikan primjer $\mathbf{x} = (2,3)$ primjenom jezgre $\kappa(\mathbf{x},\mathbf{z}) = (\mathbf{x}^\mathrm{T}\mathbf{z}+1)^2$?

   (c) Kada će broj parametara neparametarske inačice ovog modela za $n = 2$ biti veći od broja parametara njegove parametarske inačice? (U oba slučaja, parametri su vektori realnih brojeva.)

   (d) Provjerite je li u dobivenom prostoru značajki XOR-problem linearno odvojiv. Objasnite. Vrijedi li isti zaključak za jezgrenu funkciju $\kappa(\mathbf{x},\mathbf{z}) = (\mathbf{x}^\mathrm{T}\mathbf{z})^2$?

4. [***Svrha: Izvježbati izračun predikcije pomoću jezgrenog trika.***] Veza između primarnih i dualnih parametara SVM-a jest $\mathbf{w} = \sum_{i=1}^{N} \alpha_i y^{(i)} \boldsymbol{\phi}(\mathbf{x}^{(i)})$. Na skupu za učenje trenirali smo SVM s polinomijalnom jezgrenom funkcijom, $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^{\mathrm{T}}\mathbf{z} + 1)^3$. Potporni vektori su $\mathbf{x}^{(1)} = (-2, 3, 5)$, $\mathbf{x}^{(2)} = (6, 4, 3)$ i $\mathbf{x}^{(3)} = (8, 8, 2)$. Prvi primjer je negativan, a druga dva su pozitivna. Lagrangeovi koeficijenti su $\alpha_1 = 0.131$, $\alpha_2 = 0.048$ i $\alpha_3 = 0.013$. Pomak je $w_0 = -0.51$. Iskoristite jezgreni trik te odredite klasifikaciju primjera $\mathbf{x}^{(4)} = (1, 2, 3)$.

5. [***Svrha: Razumjeti karakteristike Gaussove jezgre.***]

   (a) Primjenom operacija za izgradnju složenijih Mercerovih jezgri iz jednostavnijih Mercerovih jezgri, dokažite da je Gaussova jezgra Mercerova jezgra. (Pomoć: raspišite izraz $\|\mathbf{x} - \mathbf{x}'\|^2$.)

   (b) Kako parametar $\gamma = 1/2\sigma^2$ Gaussove jezgre utječe na složenost modela? Koji je odnos između hiperparametra C i hiperparametra γ kod SVM-a?

   (c) Skicirajte očekivana područja prenaučenosti i podnaučenosti modela SVM u prostoru hiperparametara $C \times \gamma$.

   (d) Koristimo Gaussovu jezgru uz $\gamma = 1$. Možemo li u ovom slučaju odrediti u koji vektor $\boldsymbol{\phi}(\mathbf{x})$ u prostoru značajki će biti preslikan primjer $\mathbf{x}$? Možemo li odrediti težine $\mathbf{w}$. Zašto?

   (e*) Pročitajte ovo, ovo i ovo. Odgovorite: jamči li uporaba Gaussove jezgre (1) da će primjeri biti preslikani u beskonačnodimenzijski prostor značajki, (2) savršenu linearnu odvojivost primjera za učenje u prostoru značajki, (3) empirijsku pogrešku jednaku nuli na skupu za učenje, (4) minimalnu pogrešku na ispitnome skupu? Obrazložite odgovore.

6*. [***Svrha: Razumjeti na koji se način može kernelizirati algoritam linearne regresije.***] Pročitajte poglavlje 14.4.3 iz MLPP (str. 492) te izvedite kerneliziranu inačicu linearne regresije. Koja je prednost takve formulacije algoritma linearne regresije?

2 Zadatci s ispita

1. (P) Na 1000 primjera sa 100 značajki treniramo rijetki jezgreni stroj s Gaussovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 28 prototipa. **Koliko ovaj model ima hiperparametara, koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

   [A] Model nema hiperparametara, optimiramo 1001 parametara, a naučeni model ima 2857 parametara

   [B] Model ima 2800 hiperparametara, optimiramo 101 parametar, a naučeni model ima 29 parametara

   [C] Model ima 1 hiperparametar, optimiramo 1001 parametar, a naučeni model ima 2829 parametara

   [D] Model 100 hiperparametara, optimiramo 2800 parametara, a naučeni model ima 2801 parametar

2. (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

$$\mathcal{D} = \left\{(\mathbf{x}^{(i)}, y^{(i)})\right\} = \left\{((1,1),1), ((3,1),0), ((2,3),0), ((3,4),0)\right\}$$

   Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

   [A] 0 [B] 1/4 [C] 1/2 [D] 3/4

3. (N) Rješavamo problem određivanja podrijetla pojedinih riječi u jeziku: za svaku riječ trebamo odrediti je li engleskog ($y = 1$) ili francuskog ($y = 0$) podrijetla. Problem rješavamo logističkom regresijom izvedenom kao rijetki jezgreni stroj, gdje za bazne funkcije koristimo jezgru κ nad znakovnim nizovima. Funkcija κ definirana je kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Na primjer, $\kappa(\text{water}, \text{eau}) = 2/6 = 0.33$. Skup za učenje je sljedeći:

$$\begin{aligned}\mathcal{D} =& \{(\mathbf{x}, y)\}_i\\ =& \big\{(\text{water}, 1), (\text{eau}, 0), (\text{dog}, 1), (\text{chien}, 0), (\text{paperclip}, 1), (\text{trombone}), 0), (\text{chance}, 1), (\text{hasard}, 0)\big\}\end{aligned}$$

Treniranjem rijetkoga jezgrenog stroja dobili smo vektor težina $\mathbf{w} = (0.5, 0, 0, 0, -3.5, 0, 1, 0, -1)$. Razmotrite primjer $(\mathbf{x}, y) = (\text{nounours}, 0)$. **Koliko iznosi gubitak modela na primjeru $(\mathbf{x}, y)$?**

$\boxed{\text{A}}$ 0.359 $\boxed{\text{B}}$ 0.456 $\boxed{\text{C}}$ 0.552 $\boxed{\text{D}}$ 0.795

4. (N) Treniramo SVM s polinomijalnom jezgrom definiranom kao:

$$\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z})^2$$

Ova jezgra je Mercerova jezgra, što znači da postoji funkcija $\boldsymbol{\phi}$ takva da $\kappa(\mathbf{x}, \mathbf{z}) = \boldsymbol{\phi}(\mathbf{x})^T \boldsymbol{\phi}(\mathbf{z})$. Konkretno, u slučaju dvodimenzijskoga ulaznog prostora ($n = 2$), ova jezgra odgovara preslikavanju u šesterodimenzijski prostor. Međutim, postoji odstupanje u konkretnim koeficijentima polinoma. Razmotrite primjer $\mathbf{x} = (1, 0)$ te izračunajte $\boldsymbol{\phi}_\kappa(\mathbf{x})$, koji dobivamo preslikavanjem definiranim implicitno preko jezgre, te $\boldsymbol{\phi}_p(\mathbf{x})$, koji dobivamo preslikavanjem definiranim kao polinom drugog stupnja. **Koliko iznosi euklidska udaljenost između $\boldsymbol{\phi}_\kappa(\mathbf{x})$ i $\boldsymbol{\phi}_p(\mathbf{x})$?**

$\boxed{\text{A}}$ 0 $\boxed{\text{B}}$ $2\sqrt{2}$ $\boxed{\text{C}}$ 4 $\boxed{\text{D}}$ $\sqrt{2}$

5. (N) Na skupu primjera za učenje iz ulaznog prostora $n = 4$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z} + 2)^3$. Potporni vektori i njihove oznake su sljedeći:

$$\begin{aligned}(\mathbf{x}^{(1)}, y^{(1)}) &= ((9, 30, 21), -1)\\ (\mathbf{x}^{(2)}, y^{(2)}) &= ((-11, -26, -15), -1)\\ (\mathbf{x}^{(3)}, y^{(3)}) &= ((-1, -7, -6), +1)\end{aligned}$$

Lagrangeovi koeficijenti su $\alpha_1 = 2.214 \cdot 10^{-8}$, $\alpha_2 = 3.803 \cdot 10^{-8}$ i $\alpha_3 = 6.017 \cdot 10^{-8}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3, 0, -3)$.**

$\boxed{\text{A}}$ -2.330 $\boxed{\text{B}}$ -0.676 $\boxed{\text{C}}$ $+0.947$ $\boxed{\text{D}}$ $+1.434$

6. (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 5$ označenih primjera. Vektor oznaka je $\mathbf{y} = (+1, +1, -1, -1, +1)$. Euklidske udaljenosti između primjera dane su sljedećom matricom udaljenosti:

$$\mathbf{D} = \begin{pmatrix} 0.0 & 7.48 & 6.16 & 13.42 & 12.21\\ 7.48 & 0.0 & 12.73 & 20.1 & 14.18\\ 6.16 & 12.73 & 0.0 & 10.49 & 9.95\\ 13.42 & 20.1 & 10.49 & 0.0 & 20.02\\ 12.21 & 14.18 & 9.95 & 20.02 & 0.0 \end{pmatrix}$$

Treniranjem uz $C = 10$ i $\gamma = 0.0001$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (10, 1.052, 10, 10, 8.948)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za prvi primjer, $L(y^{(1)}, h(\mathbf{x}^{(1)}))$?**

$\boxed{\text{A}}$ 0.03 $\boxed{\text{B}}$ 0.24 $\boxed{\text{C}}$ 1.18 $\boxed{\text{D}}$ 1.64

7. (N) Pomoću SVM-a rješavamo problem binarne klasifikacije grafova. Budući da su primjeri $\mathbf{x}$ grafovi, koristimo SVM s jezgrenom funkcijom nad grafovima. Model treniramo na skupu od $N = 5$ označenih primjera, s vektorom oznaka jednakim $\mathbf{y} = (+1, +1, -1, -1, +1)$ i sa sljedećom jezgrenom matricom:

$$\mathbf{K} = \begin{pmatrix} 1.0 & 0.97 & -0.949 & -0.555 & -0.986\\ 0.97 & 1.0 & -0.844 & -0.336 & -0.917\\ -0.949 & -0.844 & 1.0 & 0.789 & 0.988\\ -0.555 & -0.336 & 0.789 & 1.0 & 0.684\\ -0.986 & -0.917 & 0.988 & 0.684 & 1.0 \end{pmatrix}$$

Treniranjem uz $C = 1$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (0, 0.754, 0.754, 1, 1)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za četvrti primjer, $L(y^{(4)}, h(\mathbf{x}^{(4)}))$?**

A 2.063 B 0.143 C 0.027 D 1.596

8. (P) Neka je $\mathcal{H}_{C,\gamma}$ model SVM-a s Gaussovom jezgrom. Hiperparametri tog modela su regularizacijski faktor C i preciznost jezgre γ. Odabir modela provodimo unakrsnom provjerom i to pretraživanjem po rešetci za sljedeće vrijednosti hiperparametara:

$$C = \{2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 1, 2^1, 2^2, 2^3, 2^4, 2^5\}$$
$$\gamma = \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3, 10^4, 10^5\}$$

Za model sa $C = 1$ i $\gamma = 1$ utvrdili smo da je prenaučen. **Koliko modela od ovih koje ćemo još ispitati će sigurno također biti prenaučeni?**

A 10 B 35 C 65 D 95

9. (P) Na skupu od $N = 1000$ primjera rješavamo problem višeklasne klasifikacije u $K = 4$ klase. Dvije klase imaju svaka po 400 primjera, a dvije svaka po 100 primjera. Razmatramo bismo li koristili SVM u shemi OVO ili SVM u shemi OVR. Model treniramo s jezgrenom funkcijom, no zbog ograničenja na raspoloživu računalnu memoriju moramo pripaziti da Gramova matrica ne postane prevelika. Prisjetite se da je Gramova matrična simetrična, pa je dovoljno pohraniti samo polovicu matrice (bez dijagonale). **Koji je u ovom slučaju najveći omjer veličine Gramove matrice za sheme OVO i OVR**?

A OVO:OVR $\approx$ 1:3 B OVO:OVR $\approx$ 1:405 C OVO:OVR $\approx$ 4:5 D OVO:OVR $\approx$ 32:50