8. Stroj potpornih vektora

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

Jan Šnajder, vježbe, v3.2

1 Zadatci za učenje

1. [***Svrha: Razumjeti izvod algoritma stroja potpornih vektora.***]

   (a) Definirajte, korak po korak, problem maksimalne margine (tvrda margina).

   (b) Definirajte problem kvadratnog programiranja, pripadnu Lagrangeovu funkciju te dualnu Lagrangeovu funkciju i pripadne uvjete KKT. Obrazložite svaki uvjet KKT.

   (c) Definirajte, korak po korak, dualni problem maksimalne margine te pripadne uvjete KKT koji vrijede u točki rješenja.

   (d) Koje su prednosti formulacije problema kao dualnoga optimizacijskog problema?

   (e) Napišite primarnu i dualnu formulaciju modela SVM.

   (f) Objasnite što su to potporni vektori i kako znamo da oni sigurno leže na rubu margine.

   (g) Objasnite potrebu za skaliranjem značajki kod dualne formulacije modela SVM.

2. [***Svrha: Isprobati izračuna modela potpornih vektora na konkretnom numeričkom primjeru i tako bolje razumjeti formule. Razumjeti povezanost primarne i dualne formulacije problema.***] Raspolažemo sljedećim primjerima za učenje:

$$
\begin{aligned}
\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \\
\big\{((0,0),-1),((2,4),-1),((4,2),-1),((6,4),+1),((6,8),+1),((8,8),+1)\big\}
\end{aligned}
$$

   (a) Skicirajte primjere u ulaznom prostoru $\mathbb{R}^2$ i granicu maksimalne margine. Napišite izraz za linearni model $h(\mathbf{x})$ koji odgovara toj granici.

   (b) Odredite širinu margine.

   (c) U ovom slučaju potporni vektori su $\mathbf{x}^{(3)} = (4,2)$ i $\mathbf{x}^{(4)} = (6,4)$. Odredite vektor Lagrangeovih koeficijenata $\boldsymbol{\alpha}$ temeljem izraza za ekspanziju težina $\mathbf{w}$ u potporne vektore.

   (d) Upoznajte se s formulom iz bilješke 20 iz skripte 8 te izračunajte pomak w_0.

   (e) Odredite klasifikaciju novog primjera $\mathbf{x}^{(7)} = (5,6)$ na temelju dualne formulacije modela.

2 Zadatci s ispita

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

$$
\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{((-1,1),-1),((-2,-1),-1),((2,-2),-1),((3,3),-1),((3,4),+1))\}
$$

   Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(3,3)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina šira od stare?**

   [A] $3\sqrt{2}$ puta [B] $2\sqrt{5}$ puta [C] $\sqrt{26}$ puta [D] $\frac{5}{2}\sqrt{3}$ puta

2. (N) Rješavamo binarni klasifikacijski problem. Raspolažemo označenim skupom primjera. Odgovarajuća matrica dizajna je sljedeća:

$$\mathbf{X} = \begin{pmatrix} 1 & 3 & 16 & -8 & -11 \\ 1 & -5 & 4 & -8 & -7 \\ 1 & 7 & -4 & 11 & 9 \\ 1 & 15 & -20 & 25 & 25 \end{pmatrix}$$

Na ovom skupu treniramo model SVM-a s tvrdom marginom i linearnom jezgrenom funkcijom (tj. bez preslikavanja u prostor značajki). Model treniramo u primarnoj formulaciji. Za rješenje maksimalne margine dobili smo ovaj vektor težina (uključivo s težinom w_0):

$$\mathbf{w} = (+0.1370, -0.0290, +0.0194, -0.0461, -0.0388)$$

Umjesto u primarnoj formulaciji, model smo mogli trenirati u dualnoj formulaciji, pa bismo umjesto vektora težina $\mathbf{w}$ dobili vektor dualnih parametara $\boldsymbol{\alpha}$, odnosno Lagrangeove multiplikatore. Prisjetite se da su vektori čiji su Lagrangeovi multiplikatori veći od nule potporni vektori. Premda to nije uvijek moguće, u ovom konkretnom slučaju dualni parametri modela mogu se izvesti iz rješenja primarnog modela. Izvedite vektor dualnih parametara $\boldsymbol{\alpha}$. **Koliko iznosi najveća vrijednost parametra u vektoru dualnih parametara $\boldsymbol{\alpha}$?** (Rezultate uspoređujte po prve tri decimale.)

[A] 0.0013 [B] 0.0024 [C] 0.0045 [D] 0.0089

3. (N) U ulaznome prostoru dimenzije $n = 3$ trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri $((2, -5, 15), -1)$, $((1, 8, -305), -1)$ i $((1, -6, 225), +1)$, a njima odgovarajući dualni koeficijenti su $\alpha_1 = 0.5$, $\alpha_2 = 0.8$ i $\alpha_3 = 0.9$. Treniranje smo proveli na skaliranim značajkama: svaku smo značajku x_j standardizirali primjenom transformacije $\frac{x_j - \mu_j}{\sigma_j}$, gdje su μ_j i σ_j srednja vrijednost odnosno varijanca značajke x_j u skupu označenih podataka $\mathcal{D}$. Parametri skaliranja su $\boldsymbol{\mu} = (15, -2, 100)$ i $\boldsymbol{\sigma} = (4, 1, 12)$. Model SVM-a koristimo za predikciju klase primjera $\mathbf{x} = (1, 2, -30)$. **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera x?**

[A] $+907.43$ [B] $+541.53$ [C] -373.22 [D] -739.13

4. (P) Skup za učenje čine sljedeći označeni primjeri:

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

Na skupu $\mathcal{D}$ treniramo logističku regresiju (LR) i stroj potpornih vektora s tvrdom marginom (SVM). Dodatno, treniramo model linearne regresije (LINR), gdje izlaz tog modela koristimo za klasifikaciju, tj. $h(\mathbf{x}) = \mathbf{1}\{\mathbf{w}^{\mathrm{T}}\mathbf{x} \geqslant 0\}$. Za modele SVM i LINR umjesto oznake $y = 0$ koristimo oznaku $y = -1$. Za treniranje modela LR koristimo dovoljan broj iteracija tako da možemo pretpostaviti da je dobivena pogreška unakrsne entropije praktički jednaka nuli. Razmotrite primjer $\mathbf{x}^{(7)} = (2, 0)$. Neka je $d(m)$ udaljenost primjera $\mathbf{x}^{(7)}$ od granice između klasa dobivene modelom m. **Što od navedenog vrijedi za tu udaljenost?**

[A] $d(\text{SVM}) < d(\text{LR}) < d(\text{LINR})$

[B] $d(\text{SVM}) < d(\text{LINR}) < d(\text{LR})$

[C] $d(\text{LR}) < d(\text{SVM}) < d(\text{LINR})$

[D] $d(\text{LINR}) < d(\text{LR}) < d(\text{SVM})$