8. Stroj potpornih vektora

Strojno učenje 1, UNIZG FER, ak. god. 2022./2023.

Jan Šnajder, vježbe, v3.1

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 točci 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:

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

   (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. (T) Kod izvoda algoritma SVM s tvrdom marginom, pretpostavili smo da za skup označenih primjera $\mathcal{D}$ vrijedi sljedeći uvjet linearne odvojivosti:

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

Koliko hipoteza zadovoljava ovaj uvjet, i kako algoritam SVM odabire jednu od njih?

A Uvjet zadovoljava konačan broj hipoteza koje su linearno odvojive, a SVM između njih odabire onu jednu koja minimizira zbroj kvadrata težina

B Uvjet zadovoljava konačan broj hipoteza koje su linearno odvojive, no one se razlikuju samo po faktoru koji množi težine $(\mathbf{w}, w_0)$, pa SVM odabire onu jednu za koju vrijedi $yh(\mathbf{x}) \geqslant 1$ za sve primjere

C Uvjet zadovoljava beskonačno mnogo hipoteza, a SVM odabire onu jednu koja minimizira zbroj kvadrata težina te koja ispravno klasificira sve primjere, uz uvjet da $h(\mathbf{x})$ nije u intervalu $(-1, +1)$

D Uvjet zadovoljava beskonačno mnogo hipoteza, međutim samo za jednu vrijedi $yh(\mathbf{x}) = 1$ za najbliže primjere, i to je hipoteza koju odabire SVM

2. (T) Kod SVM-a, problem maksimalne margine sveo se na problem minimizacije izraza $\frac{1}{2}\|\mathbf{w}\|^2$ uz određena ograničenja. **Zašto minimizacija ovog izraza daje maksimalnu marginu?**

A Što je vektor $\mathbf{w}$ kraći, to je manja udaljenost d primjera od hiperravnine, a to efektivno znači da je margina to šira jer je margina fiksna a udaljenosti d se smanjuju

B Što je vektor $\mathbf{w}$ kraći, to je manja vrijednost $h(\mathbf{x})$, ali je težina w_0 konstantna, pa se udaljenosti između hiperravnine i primjera povećavaju, što znači da se margina širi

C Što je vektor $\mathbf{w}$ kraći, to je manja vrijednost $h(\mathbf{x})$, pa primjeri moraju biti što dalje da bi vrijedilo $h(\mathbf{x}) = \pm 1$, a to znači da je margina to šira

D Što je vektor $\mathbf{w}$ kraći, to je veća udaljenost d primjera od hiperravnine, što znači da se potporni vektori udaljavaju od hiperravnine, a to znači da margina postaje šira

3. (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Bez induktivne pristranosti nije moguće naučiti model koji bi generalizirao. **Po čemu se induktivna pristranost algoritma SVM (tvrda margina) razlikuje od induktivne pristranosti algoritma perceptrona?**

A Razlikuju se po pristranost preferencijom, jer perceptron ne maksimizira marginu, premda se može dogoditi da pronađe rješenje koje maksimizira marginu

B SVM ima pristranost preferencijom kojom maksimizira marginu, dok perceptron nema induktivnu pristranost preferencijom već samo pristranost jezika

C Imaju istu pristranost preferencijom, a to je da primjeri moraju biti linearno odvojivi, no SVM ima dodatnu pristranost ograničenjem u vidu optimizacijskih ograničenja

D Imaju istu pristranost jezika, a pristranost preferencijom također će biti ista ako se oba optimiraju gradijentnim spustom s istim početnim težinama i istom stopom učenja

4. (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

5. (T) Kod optimizacije SVM-a iskoristili smo Lagrangeovu dualnost kako bismo se iz primarnog optimizacijskog problema prebacili u dualni optimizacijski problem. To smo učinili tako da smo na temelju Lagrangeove funkcije L definirali dualnu Lagrangeovu funkciju $\tilde{L}$ i uveli nova ograničenja, što nam je opet dalo kvadratni program. **Kako onda u konačnici glasi optimizacijski problem tvrde margine u dualnoj formulaciji (ako zanemarimo ograničenja)?**

A $\operatorname{argmax}_{\boldsymbol{\alpha}} \min_{\mathbf{w},w_0} L(\mathbf{w}, w_0, \boldsymbol{\alpha})$

B $\operatorname{argmin}_{\boldsymbol{\alpha}} \max_{\mathbf{w},w_0} L(\mathbf{w}, w_0, \boldsymbol{\alpha})$

C $\operatorname{argmax}_{\mathbf{w},w_0} \min_{\boldsymbol{\alpha}} L(\mathbf{w}, w_0, \boldsymbol{\alpha})$

D $\operatorname{argmin}_{\mathbf{w},w_0} \max_{\boldsymbol{\alpha}} L(\mathbf{w}, w_0, \boldsymbol{\alpha})$

6. (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

7. (T) Model SVM-a može se definirati i optimirati u primarnoj ili dualnoj formulaciji. **Konceptualno, kada će primjer x u dualnoj formulaciji SVM-a biti klasificiran u pozitivnu klasu?**

A Ako je linearna kombinacija značajki iz $\mathbf{x}$ s pozitivnim težinama veća ili jednaka linearnoj kombinaciji značajki iz $\mathbf{x}$ s negativnim težinama

B Ako je vektor $\mathbf{x}$ po skalarnom umnošku sličniji potpornim vektorima s pozitivnom oznakom nego potpornim vektorima s negativnom oznakom

C Ako je skalarni umnožak vektora $\mathbf{x}$ i vektora oznaka $\mathbf{y}$ veća od praga definiranog parametrom w_0

D Ako većina od ukupno α primjera iz skupa za učenje koji su po euklidskoj udaljenosti najbliži primjeru $\mathbf{x}$ ima pozitivnu oznaku