

9. Stroj potpornih vektora II

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

Jan Šnajder, vježbe, v2.7

1 Zadatci za učenje

- [Svrha: Razumjeti potrebu za mekom marginom. Znati izvesti problem meke margine SVM-a preko Lagrangeove dualnosti.]**
 - Objasnite motivaciju za uvođenje meke margine. Skicirajte primjer prenaučenosti kod tvrde margine, i to za linearno odvojiv i linearno neodvojiv slučaj.
 - Formulirajte problem optimizacije meke margine.
 - Definirajte dualni kvadratni problem za meku marginu.
 - Krenuvši od uvjeta KKT, dokažite da potporni vektori za koje vrijedi $0 < \alpha_i < C$ leže na margini, a da vektori za koje $\alpha_i = C$ leže na margini ili se nalaze unutar nje.
- [Svrha: Znati izvesti formulaciju algoritma SVM preko gubitka zglobnice. Razumijeti funkciju pogreške SVM-a.]**
 - Krenuvši od problema meke margine, izvedite gubitak zglobnice.
 - Napišite empirijsku pogrešku SVM-a i izrazite vezu između hiperparametara C i regularizacijskog faktora λ .
 - Razmotrite zadatak 2 iz vježbi 8. Pretpostavite da je ispravna klasifikacija primjera $\mathbf{x}^{(7)} = (5, 6)$ iz (e) dijela zadatka negativna. Koliko iznosi gubitak koji primjer $\mathbf{x}^{(7)}$ nanosi SVM modelu iz tog zadatka?
 - Skicirajte pogrešku učenja i pogrešku ispitivanja kao funkciju od C . Kojem području odgovara prenaučenost a kojem podnaučenost?
- [Svrha: Razumjeti kako se gubitak zglobnice razlikuje od ostalih funkcija gubitaka koje smo razmatrali. Razumjeti kako gubitci određuju robusnost klasifikacijske granice.]** Raspolažemo sljedećim primjerima za učenje:

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

- Skicirajte funkcije gubitka L kao funkcije od $\mathbf{y}\mathbf{w}^T\phi(\mathbf{x})$ za gubitak (1) linearne regresije, (2) perceptrona, (3) logističke regresije i (4) stroja potpornih vektora.
- Pozivajući se na skice funkcija gubitka, skicirajte predvidive hipoteze ova četiri algoritma.
- Načinite skicu kao za prethodni zadatak, ali za skup podataka u koji je dodan primjer $((8, 1), 1)$. Komentirajte razliku u odnosu na prethodnu skicu.
- Pokušajte odgovoriti: zašto algoritam SVM-a često daje rijetke modele, unatoč tome što zapravo koristi L2-regularizaciju, za koju je poznato da ne rezultira rijetkim modelima? (Pomoć: usporedite gubitak zglobnice i gubitak logističke regresije.)

2 Zadatci s ispita

1. (T) Problem meke margine SVM-a formulirali smo kao problema optimizacije uz ograničenja, preciznije kao problem kvadratnog programiranja. Neka je n broj značajki, a N broj primjera. **Koliko primarni optimizacijski problem meke margine ima ukupno ograničenja, a koliko varijabli po kojima optimiramo?**

- ☐ A N ograničenja i $2N + 1$ varijabli
☐ B N ograničenja i $n + 1$ varijabli
☐ C $2N$ ograničenja i $N + n + 1$ varijabli
☐ D $2N$ ograničenja i $2n$ varijabli

2. (T) Kod optimizacijskog problema meke margine jedan od uvjeta KKT koji vrijede u točki rješenja je sljedeći uvjet komplementarne labavosti:

$$\alpha_i (y^{(i)} (\mathbf{w}^T \mathbf{x}^{(i)} + w_0) - 1 + \xi_i) = 0$$

Što možemo zaključiti na temelju ovog uvjeta?

- ☐ A Primjeri koji nisu potporni vektori sigurno se nalaze izvan margine
☐ B Potporni vektori nisu izvan margine i na pravoj su strani granice
☐ C Potporni vektori nalaze se na margini ili izvan nje, a na pravoj strani granice
☐ D Primjeri koji nisu potporni vektori nalaze se na margini ili unutar margine

3. (N) Raspolažemo sljedećim skupom označenih primjera u trodimenzijskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{((-1, 3, 6), -1), ((-4, 4, 4), -1), ((-2, 4, 1), +1)\}$$

Na ovom skupu primjera treniramo model SVM-a s linearnom jezgrenom funkcijom i sa $C = 0.01$. Postupak treniranja algoritmom SMO završio je s vektorom Lagrangeovih koeficijenata $\boldsymbol{\alpha} = (0, 0.01, 0.01)$. Iz ovoga se može izračunati da vrijedi $w_0 = -0.8$. Umjesto algoritma SMO, za optimizaciju smo mogli upotrijebiti gradijentni spust i optimirati težine u primarnoj formulaciji problema. U tom slučaju koristili bismo empirijsku pogrešku SVM-a definiranu kao L2-regularizirani gubitak zglobnice. Međutim, tu pogrešku možemo izračunati i naknadno, nakon što smo naučili model. **Koliko iznosi empirijska pogreška ovog SVM-a na skupu primjera \mathcal{D} ?**

- ☐ A 1.935 ☐ B 33.935 ☐ C 1.135 ☐ D 33.135