Izveštaj o trećem domaćem zadatku iz predmeta Mašinsko učenje

Viktor Todosijević 3140/2021

Decembar 2022

1 Uvod

Tema drugog domaćeg zadatka je bila "Metoda nosećih vektora". Student je u programskom jeziku *Python* implementirao obučavanje i primenu metode nosećih vektora. U odeljku 2 će biti reči o samoj metodi, dok će u odeljku 3 biti reči o implementaciji, a u 4 o rezultatima na skupu podataka priloženom uz zadatak.

2 Metoda nosećih vektora

Metoda nosećih vektora nastoji da projektuje klasifikator odbiraka na takav način da u opštem slučaju maksimizuje razdaljinu(marginu) izmedju odbiraka klasa i diskriminacione prave. Na ovaj način klasifikator izbegava preobučavanje i maksimizuje mogućnost generalizovanja na nevidjene odbirke jer su obučavajući odbirci u načelu udaljeni od diskriminacione prave, stoga ako je nevidjeni odbirak iz iste raspodele kao i obučavajući, ali malo pomeren, verovatno je da neće preći diskriminacionu krivu. U slučaju da je nemoguće rastaviti dve klase linearnim klasifikatorom uvodimo pojam šarka-gubitaka u optimizacionom problemu koje dozvoljavaju da pojedini odbirci ne budu veoma udaljeni od diskriminacione prave. Obučavanje ovakvog klasifikatora se svodi na sledeći optmizacioni problem [1]:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\boldsymbol{\omega}^T\boldsymbol{\omega} + C\sum_{n=1}^{N}\xi_n \\ \text{t.d.} \quad & t_n y(\boldsymbol{x}_n) + \xi_n - 1 \geq 0 \\ & \xi_n \geq 0 \\ & \forall\ n \in [1, N] \end{aligned} \tag{1}$$

gde je $\boldsymbol{\omega}$ vektor koeficijenata klasifikatora, b njegov pomeraj, C hiperparametar, ξ vektor šarka-gubitaka, N broj odbiraka u obučavajućem skupu, t_n oznaka klase odbirka iz obučavajućeg skupa, a $\boldsymbol{x}_n$ odbirak iz obučavajućeg skupa. Kako je ovaj problem konveksan možemo primeniti metod lagranžijana i KKT uslove kako bi dobili primalnu 2 i dualnu 3 formulaciju problema:

$$L_p = \frac{1}{2}||\boldsymbol{\omega}||^2 + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N}a_n(t_n y(\boldsymbol{x}_n) - 1 + \xi_n) - \sum_{n=1}^{N}\mu_n\xi_n \tag{2}$$

$$L_d = \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m t_n t_m (\boldsymbol{x}_m^T \boldsymbol{x}_n), \quad \text{t.d.} \sum_{n=1}^{N} a_n t_n = 0 \tag{3}$$

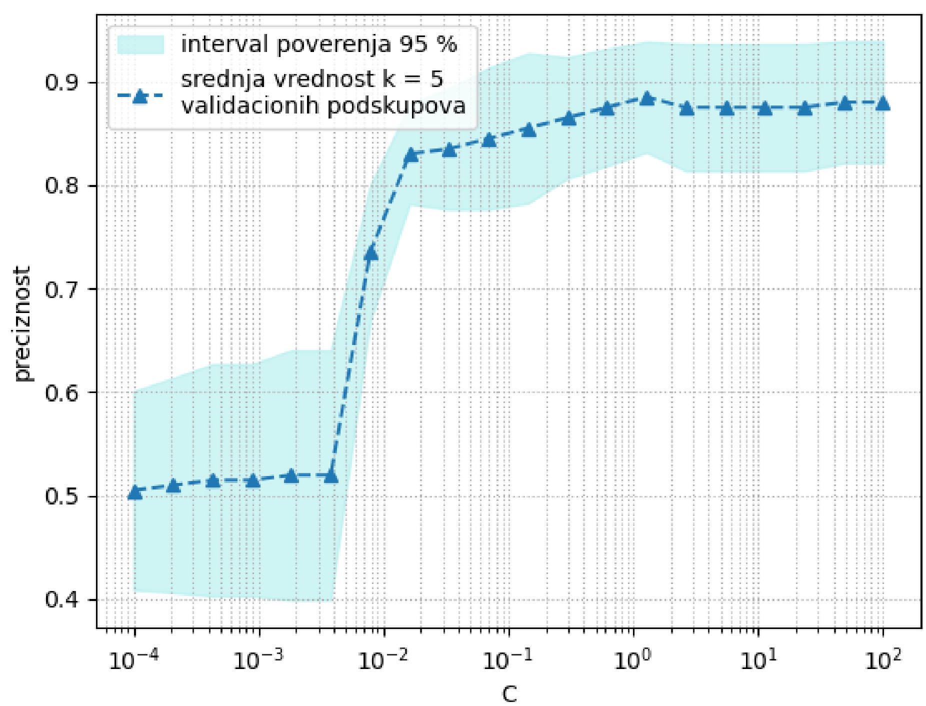
gde su a_n i μ_n Lagranževi multiplikatori za nejednakosti iz postavke problema 1 za koje važe KKT uslovi. Ako proizvod $\boldsymbol{x}_m^T \boldsymbol{x}_n$ iz 3 zamenimo kernelskom funkcijom $K(\boldsymbol{x}_m, \boldsymbol{x}_n)$ možemo dobiti i nelinearne diskriminacione funkcije.

3 Metodologija

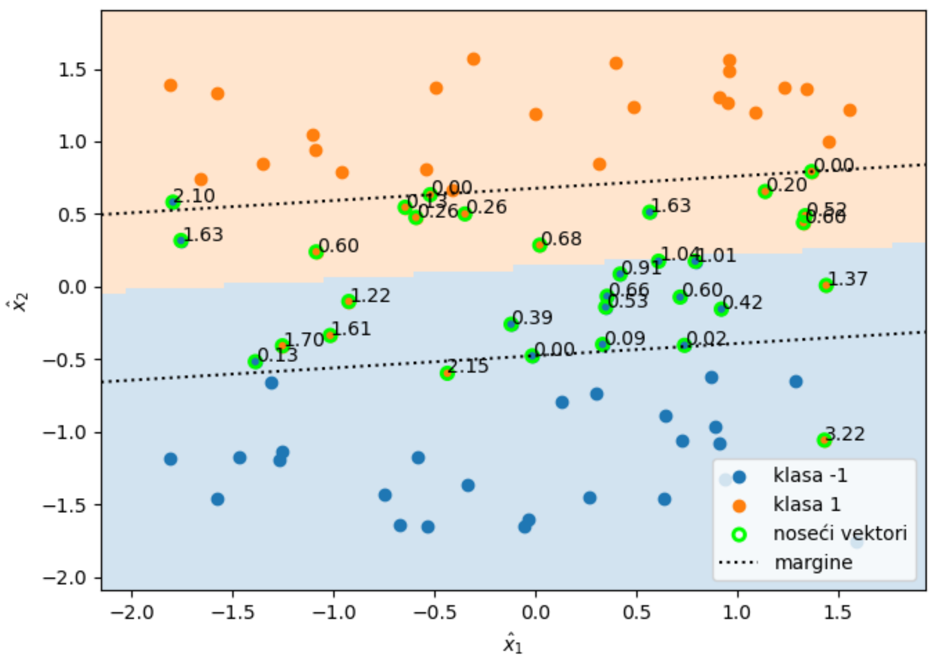
Klasifikatore obučavamo tako što pomoću programskog paketa *cvxopt* rešavamo kvadradni program ili u primalnom ili u dualnom obliku. Rešenje u primalnom obliku rezultuje linearnim klasifikatorom dok se na dualni problem može primeniti metoda kernela i dobiti nelinearni klasifikator. U ovoj implementajici koristimo Gausovu kernelsku funkciju. Pre obučavanja normalizujemo podatke. Pretragom po prostoru hiperparametara odredjujemo C i standardnu devijaciju Gausove funkcije σ primenom krosvalidacije u $k = 5$ validacionih podskupova.

4 Rezultati

Rešenje primalnog problema rezultuje prosečnom tačnošću na testirajućem skupu od 88% pri vrednosti hiperparametra $C = 1$ dobijenog pretragom prikazanom na slici 1. Način na koji klasifikator deli prostor, kao i noseći vektori, margine i šarka gubici prikazani su na slici 2.

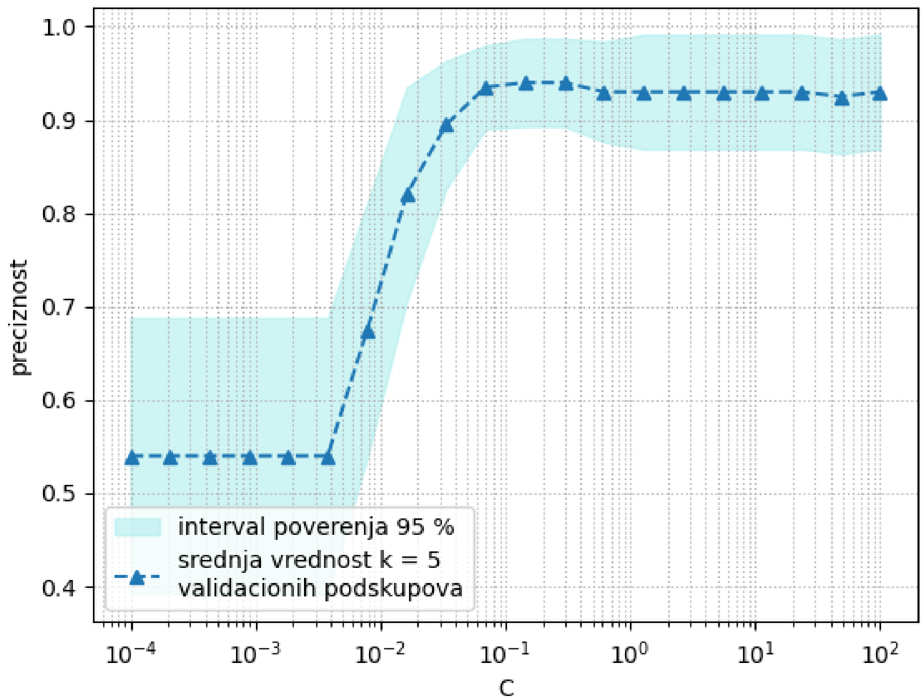


Slika 1: Pretraga po prostoru hiperpatametara C za primalni problem.

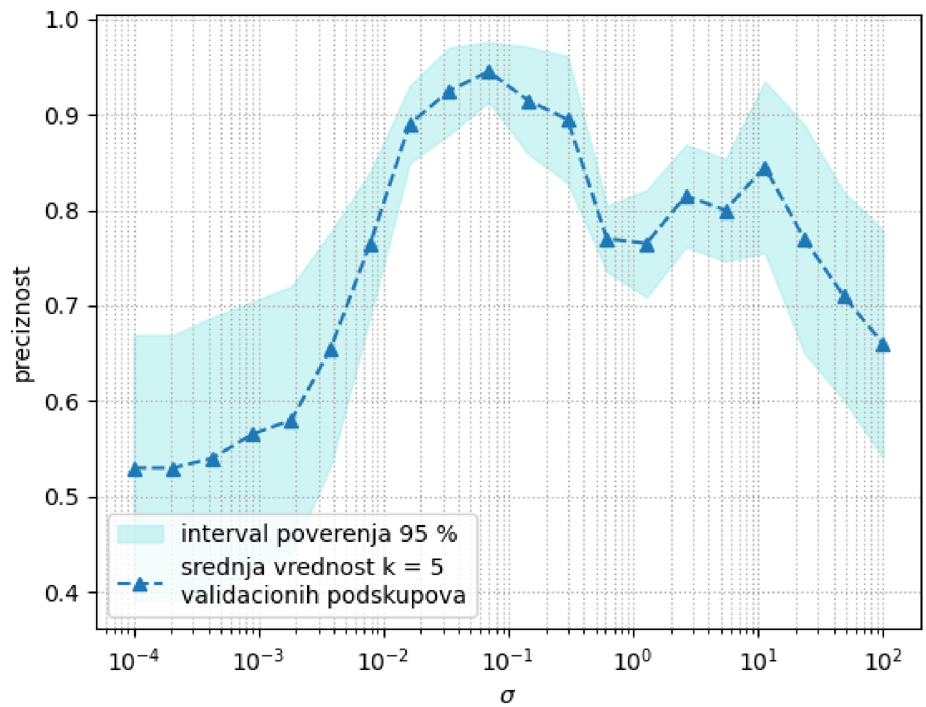


Slika 2: Prostor odbiraka podeljen od strane klasifikatora kao i margine, noseći vektori i njihovi šarka-gubici nakon rešavanja primalnog problema.

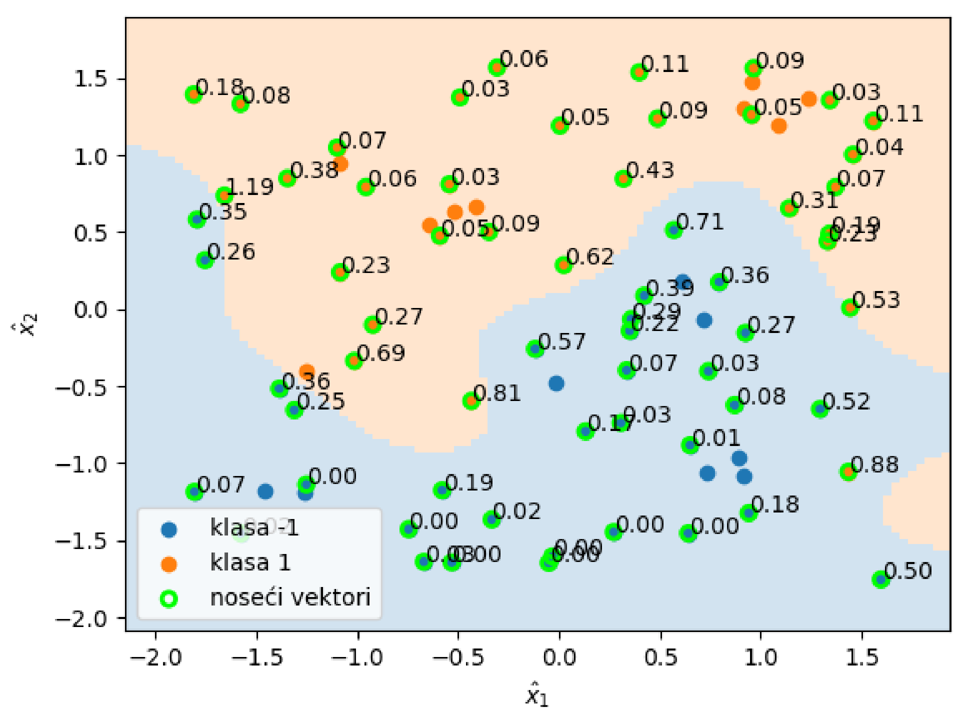
Rešenje dualnog problema rezultuje prosečnom tačnošću na testirajućem skupu od 93% pri vrednosti hiperparametra $C = 0.5$ i $\sigma = 0.1$ dobijenih pretragama prikazanim na slikama 3 i 4. Način na koji klasifikator deli prostor, kao i noseći vektori i šarka gubici prikazani su na slici 5. Na istoj slici se jasno vidi da su skoro svi odbirci noseći; student ne zna zašto je to tako.



Slika 3: Pretraga po prostoru hiperpatametara C za dualni problem pri $\sigma = 0.1$.



Slika 4: Pretraga po prostoru hiperpatametara σ za dualni problem pri $C = 0.1$.



Slika 5: Prostor odbiraka podeljen od strane klasifikatora kao i margine, noseći vektori i njihovi šarka-gubici nakon rešavanja dualnog problema.

Reference

[1] Cristopher M. Bishop. Pattern Recognition and Machine Learning, Springer, 2006