

Elektrotehnički fakultet

Izveštaj domaćeg zadatka

Prepoznavanje oblika

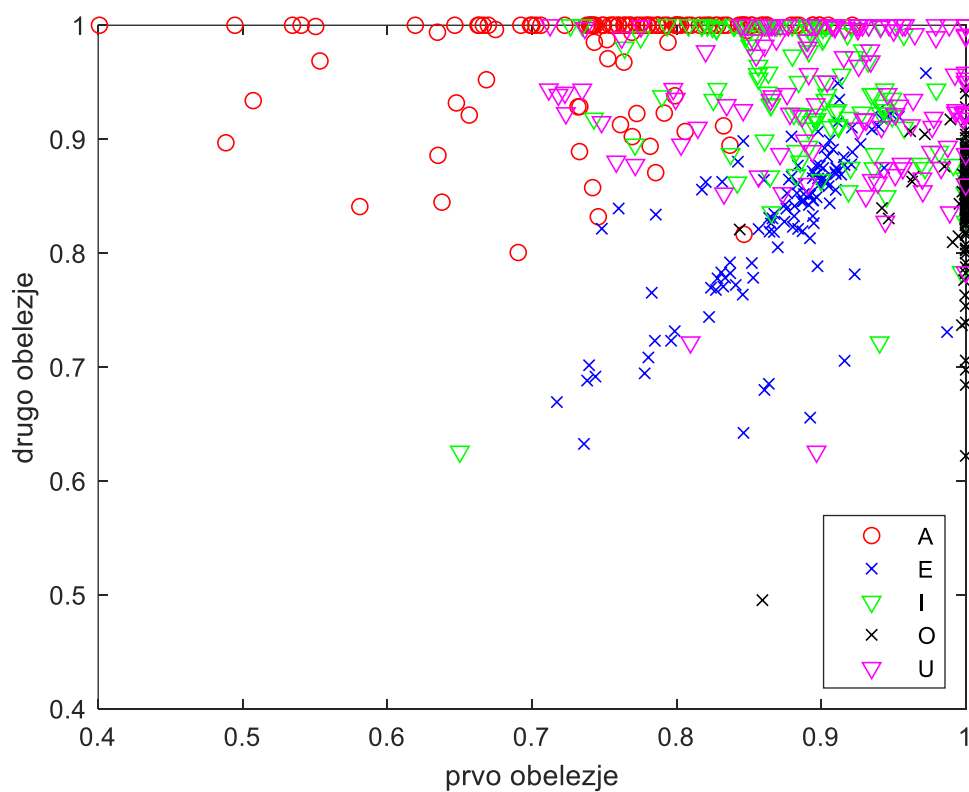
Nikolina Bunijevac

Prvi zadatak

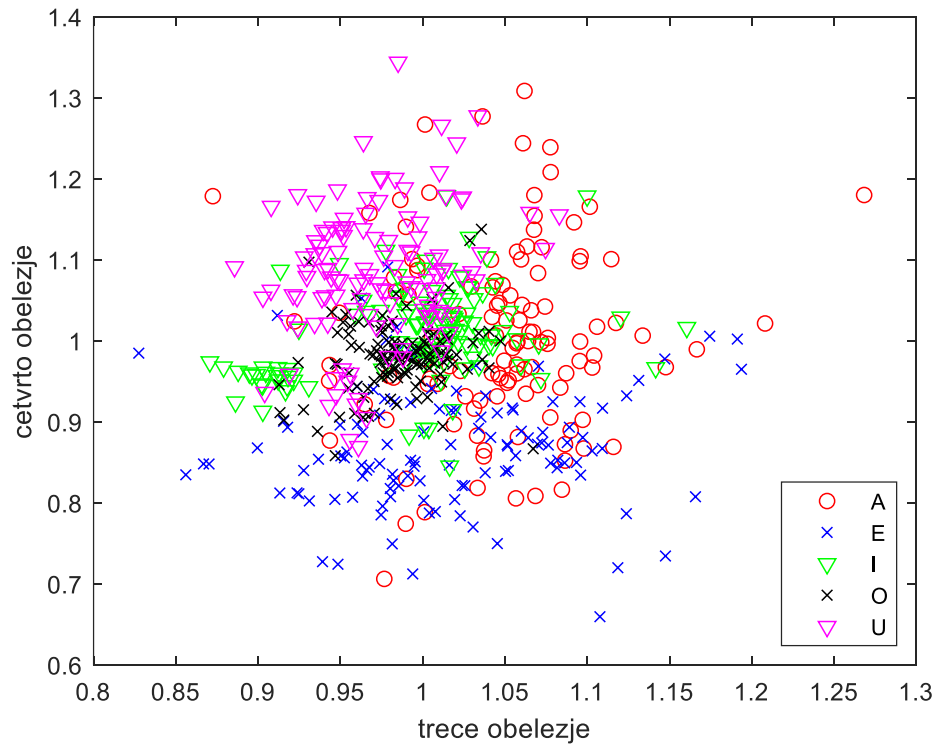
Sistem zasnovan na testiranju hipoteza koristi informaciju o sledećih pet obeležja:

1. Mean vrednost u sredini slike
2. Mean u središnjoj donjoj četvrtini
3. Odnos broja piksela u gornjoj i donjoj polovini slike
4. Odnos broja piksela u levoj i desnoj trećini slike
5. Broj preseka na 1/3 slike (pomaže da bolje razlikujemo A od I)

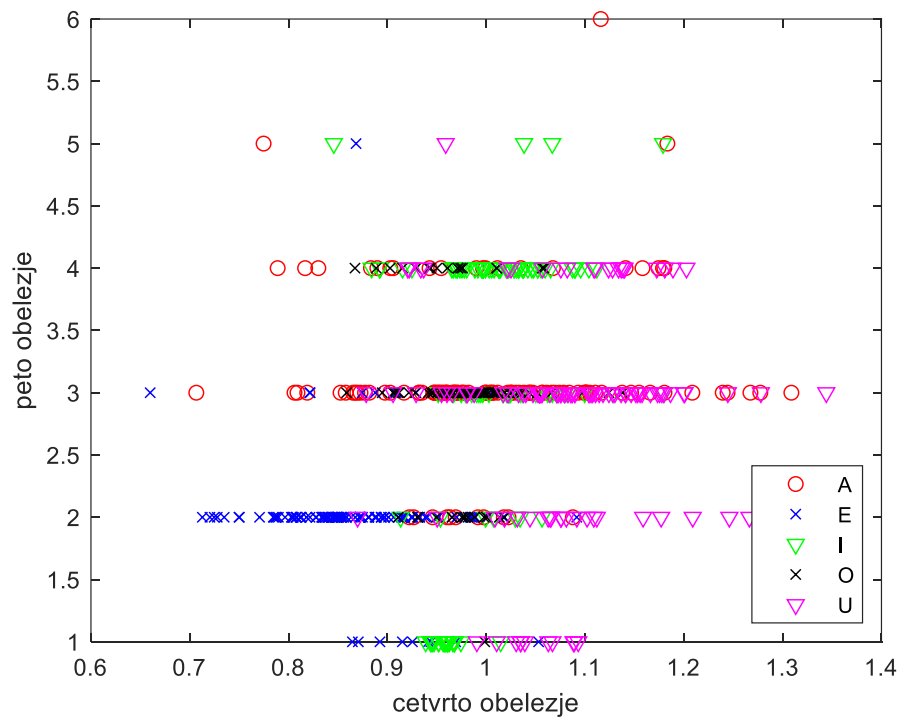
Na sledećim graphicima prikazana su obeležja po klasama, 2 po 2. Možemo videti na slikama 1.1-1.3 da su nam neka informativnija za određena slova (npr. prvo obeležje specifično dobro izdvaja slovo O).



Slika 1.1: Primećujemo da možemo da očekujemo probleme sa razlikovanjem slova E, I i U.



Slika 1.2: Sada je E izdvojenije.



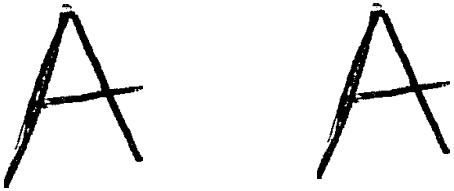

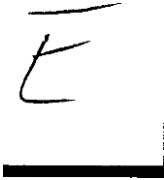


Slika 1.3: A i I imaju manje zajedničkih vrednosti za peto obeležje, pa će nam ovo pomoći da ih bolje klasifikujemo.

Matrica konfuzije:

	A	E	I	O	U
A	18	0	2	0	0
E	0	16	1	0	3
I	0	0	16	1	3
O	0	0	0	20	0
U	0	0	0	0	20

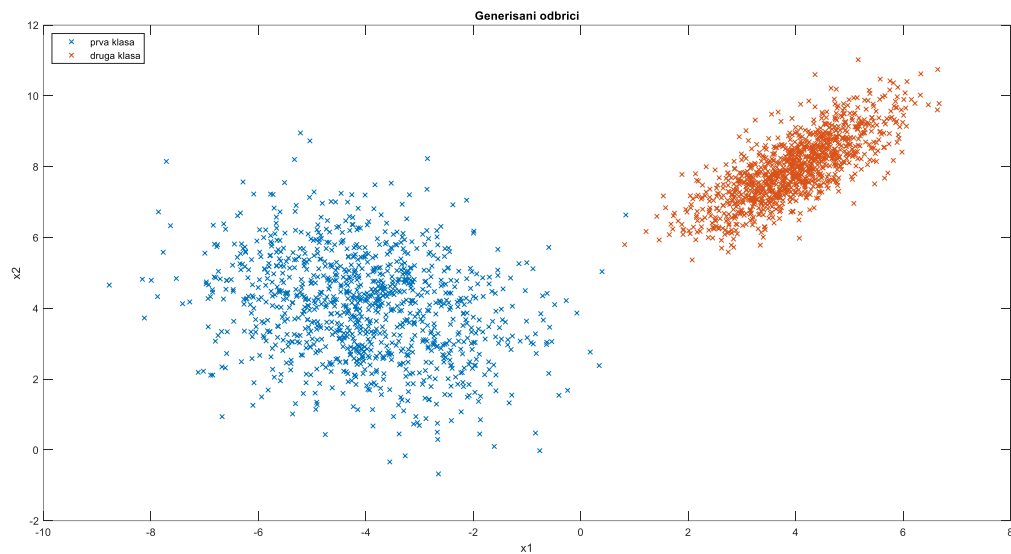
Greška klasifikacije: 0.1

Pogrešno klasifikovana slova:

A		<p>Ova slova A su klasifikovana kao slovo I. Slova su prilično slična, podebljana leva kosa linija, te je verovatno detektovano više gornjih poprečnih preseka, što je karakteristično za slovo I.</p>
E	 <p>Zajedničko za ova slova jeste duža središnja poprečna linija, što je dovelo do pomeranja „težišta” na levo, što je karakterističnije za slovo U, kako su pogrešno i klasifikovana ova slova.</p>	 <p>Loša odrađeno odsecanje, slovo klasifikovano kao I.</p>
I	 <p>Klasifikovana su kao U, takođe moguće zbog toga što su leve linije „jače” od desnih. Segmentacija je mogla biti bolje urađena.</p>	 <p>Klasifikovano kao O.</p>

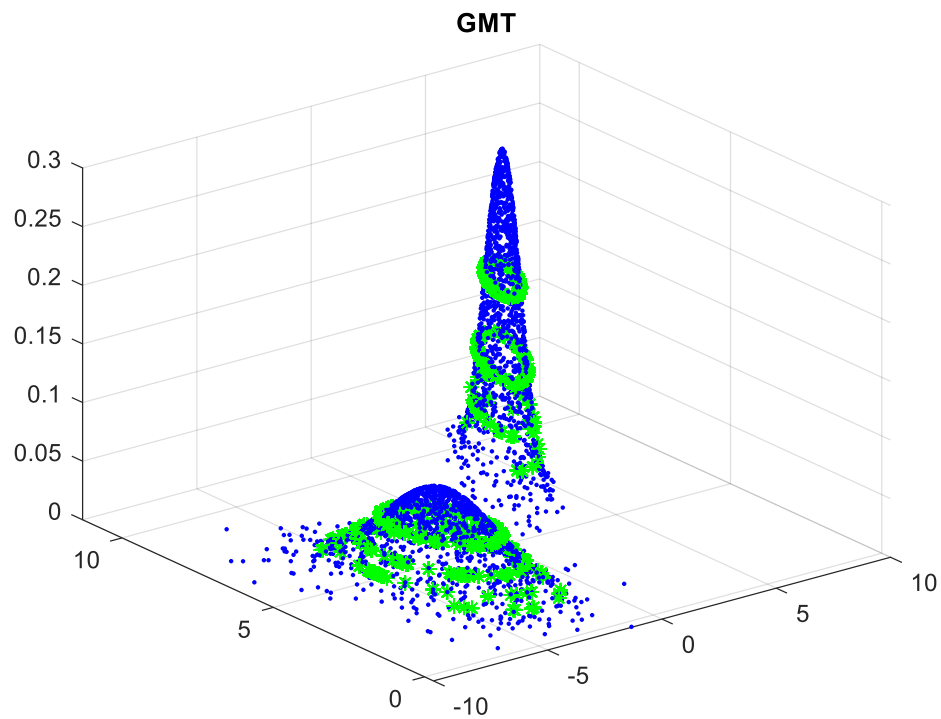
Drugi zadatak

a)



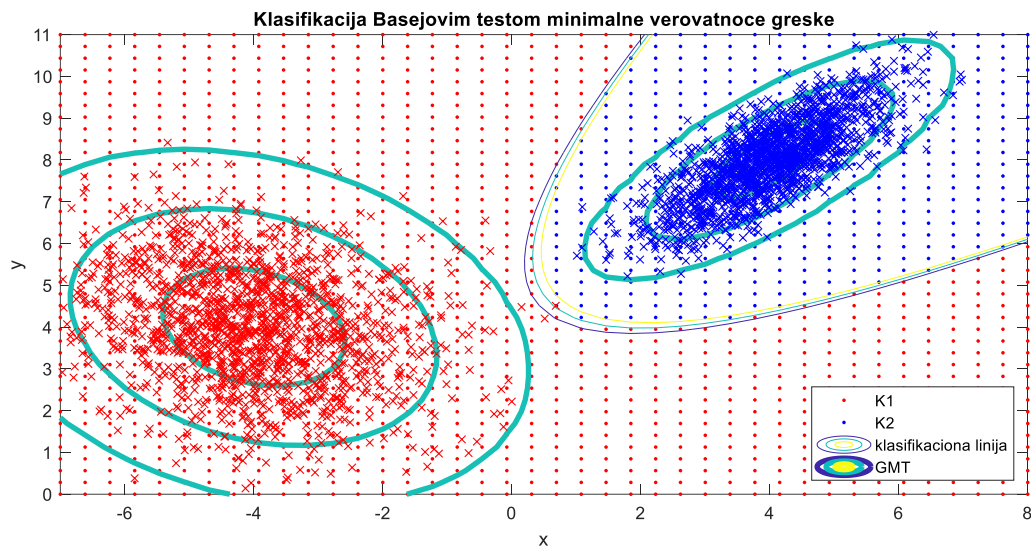
Slika 2.1. Na slici su prikazani odbirci prve i druge klase.

b)

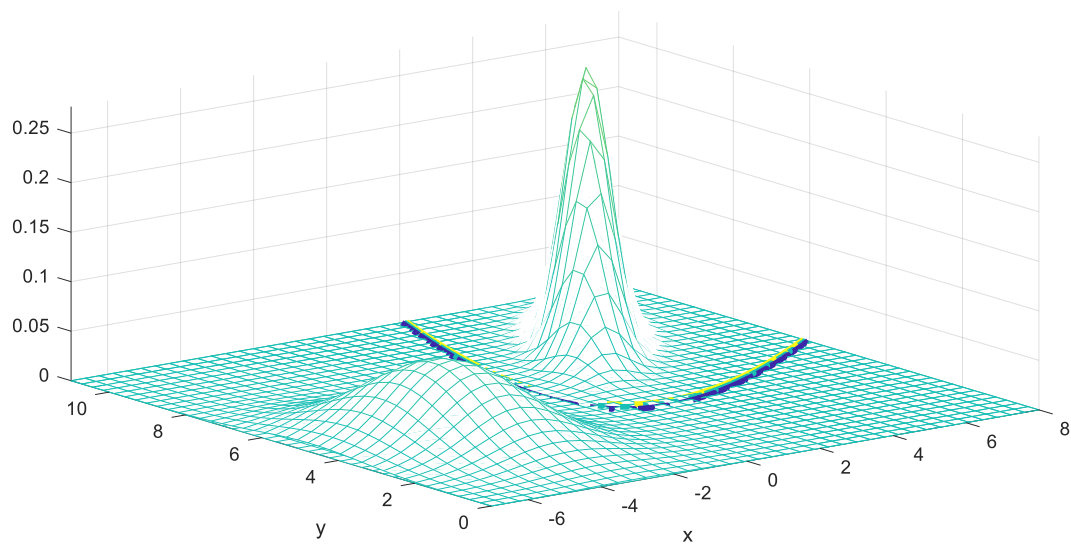


Slika 2.2: Na slici su prikazane d-kvadrat krive za odbirke dvaju klasa na dijagramu u prostoru oblika

c)

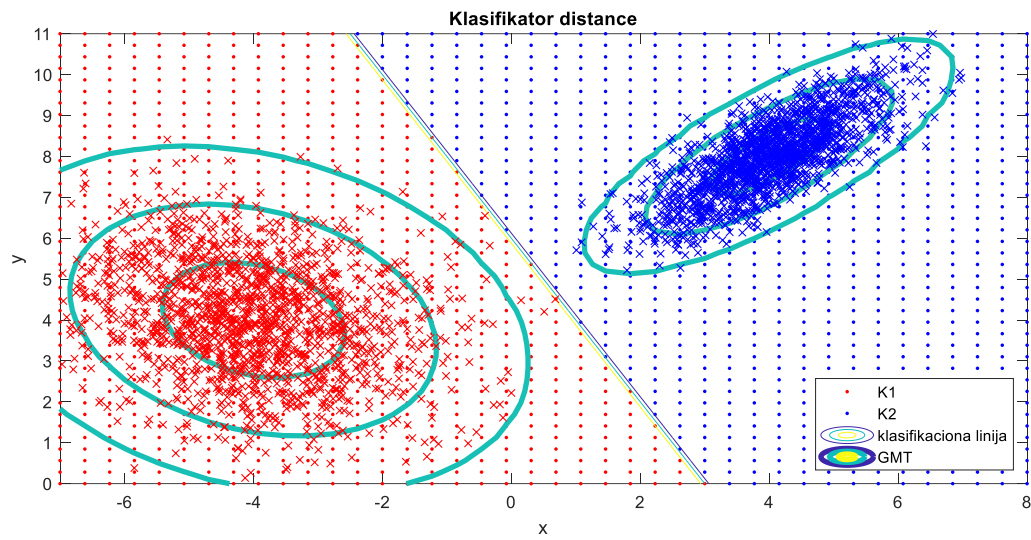


Slika 2.3: Obirci raspoređeni Bajesovim testom minimalne verovatnoće greške, sa 3 praga u zavisnosti od težine (bitnosti) klasa. Vidimo u slučaju kada je bitnija prva klasa, kada imamo žutu klasifikacionu liniju, da je svaki odabirak prve klase iz generisanog skupa podataka smešten u prvu klasu, što je očekivano. Greška klasifikacije kada su klase iste težine iznosi približno 10^{-4} za obe klase.



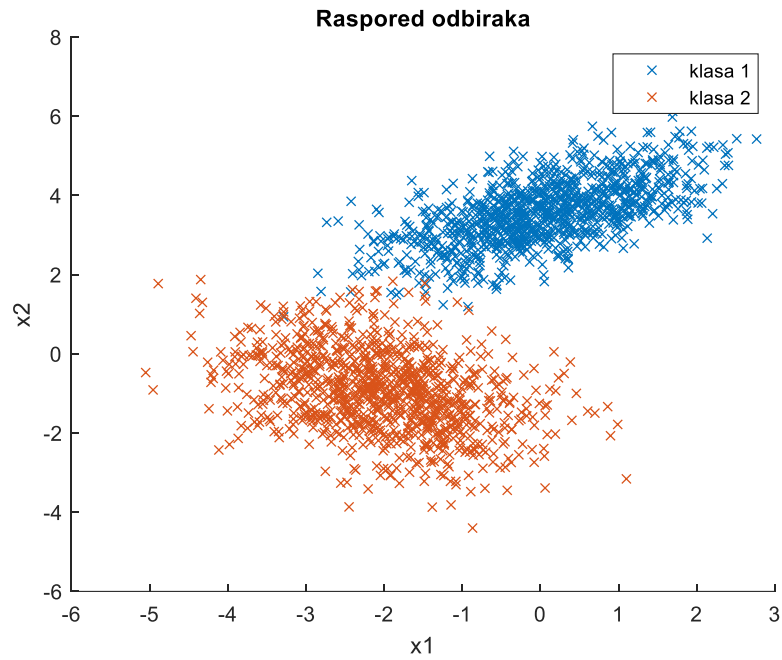
Slika 2.4: 3D prikaz slike 2.3.

d)

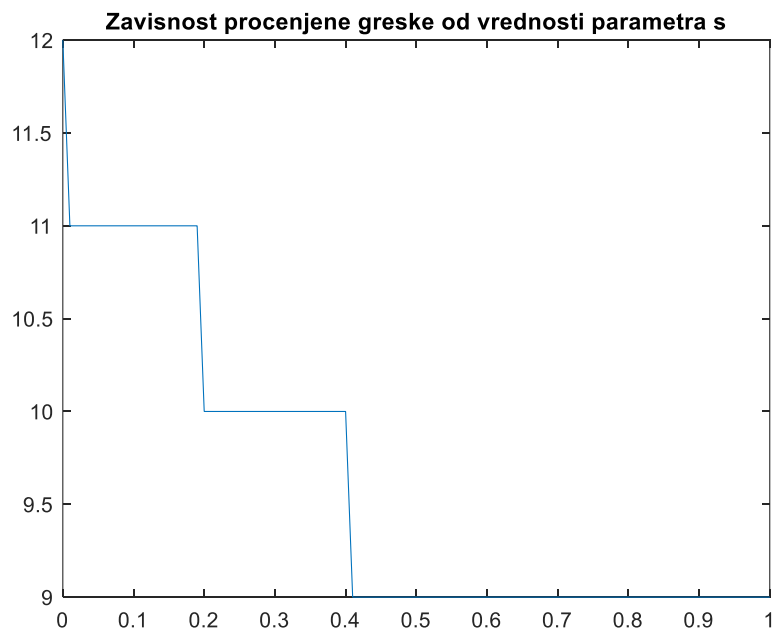


Slika 2.5: Klasifikator distance uspešno je klasifikovao odbirke klasa zato što su linearno separabilne. Greška klasifikacije su blago veće nego u prethodnom slučaju: oko 0.0002 za grešku prvog i tipa i oko 0.0001 za grešku drugog tipa.

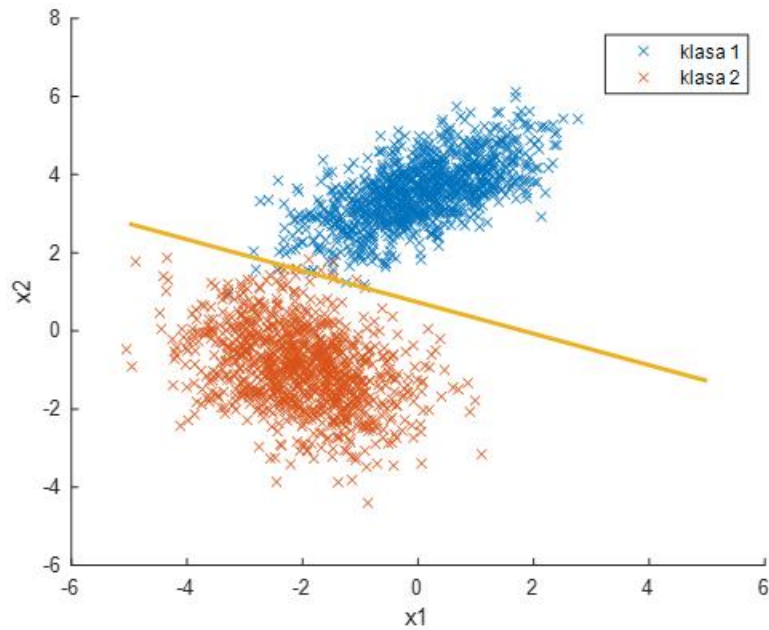
Treći zadatak



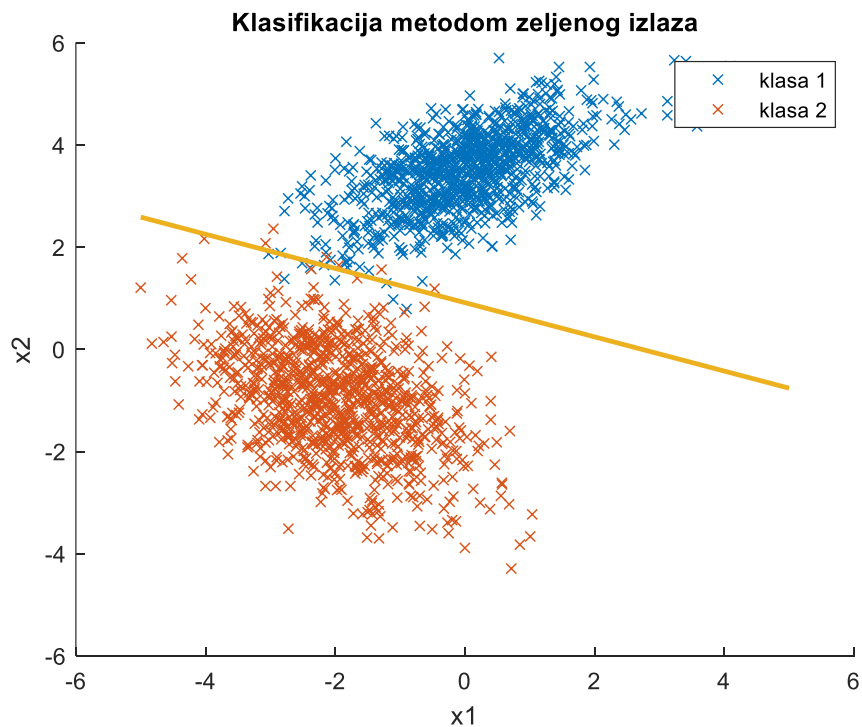
Slika 3.1: Raspored odbiraka dvaju klasa



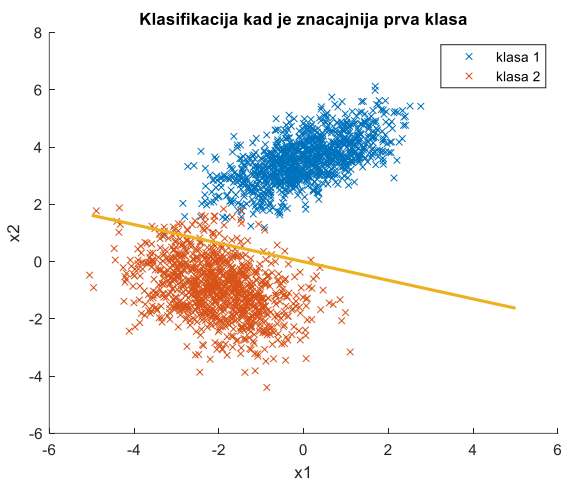
Slika 3.2: Klasifikator je projektovana drugom iterativnom metodom. Na slici je prikaz zavisnosti procenjene greške u zavisnosti od parametra s .



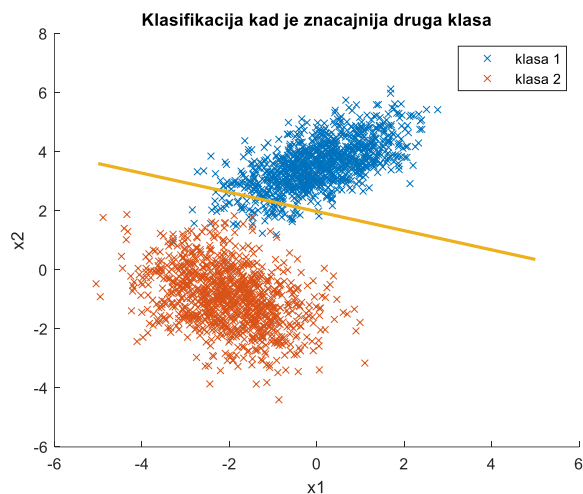
Slika 3.3: Na grafiku iznad prikazan je klasifikaciona linija na osnovu linearnog klasifikatora isprojektovanog drugom iterativnom metodom. Dobijeni parametri su: $v_0 = 2.4100$, $V = [-2.3060; -5.3874]$.



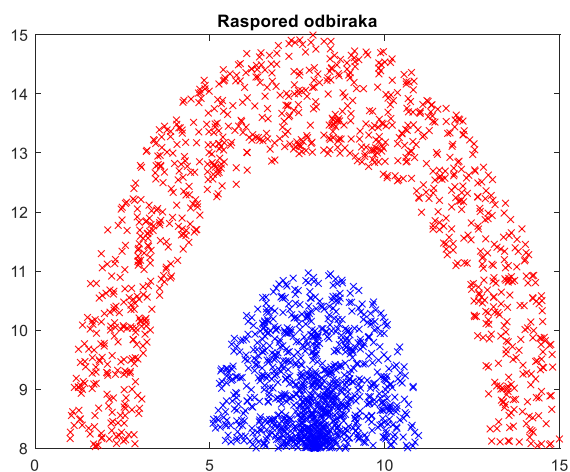
Slika 3.4: U ovom slučaju, klase su podjednako bitne, a dobijena matrica: $W = [0.3240; -0.1077; -0.3427]$



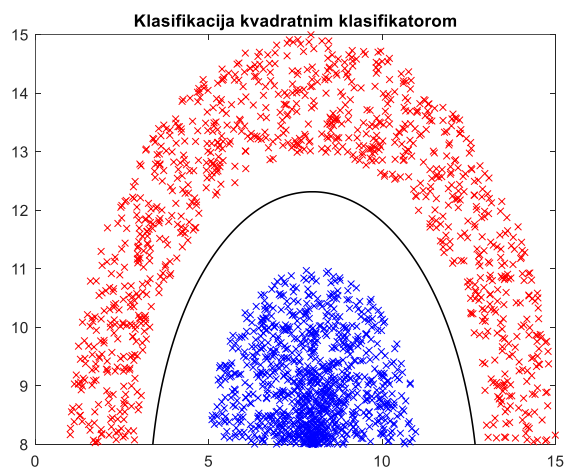
Slika 3.5: Kada je prva klasa značajnija, klasifikaciona linija se udaljava od nje, kako bi odbirci te klase bili ispravno klasifikovani.



Slika 3.6: Na isti način, kada je druga klasa značajnija, klasifikaciona linija se od nje udaljila.



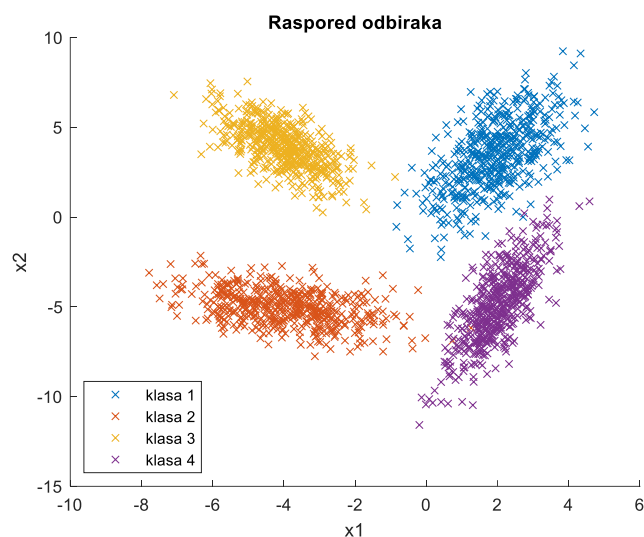
Slika 3.7: Raspored dve klase koje nisu linearno separabilne, dakle moramo primeti kvadratni klasifikator.



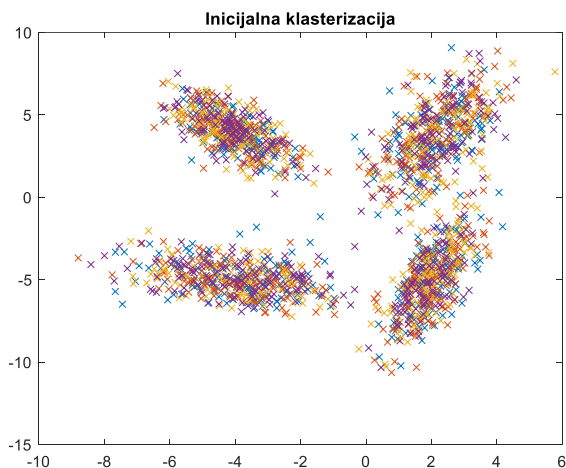
Slika 3.8: Kvadratni klasifikator sa parametrima: $Q = [0.0530, 0.0006; 0.0006, 0.0412]$

Četvrti zadatak

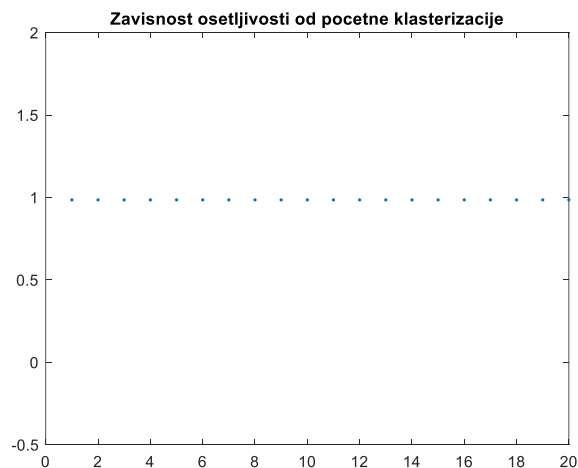
a)



Slika 4.a.1: Raspored odbiraka četiri linearno separabilne klase. Pokušaćemo klasifikaciju da izvršimo cmean metodom, za koju očekujemo da bude uspešna jer su klase linearno separabilne, a ona po svojoj prirodi treba da reši probleme tog tipa (klasifikacione linije su bisektrise).



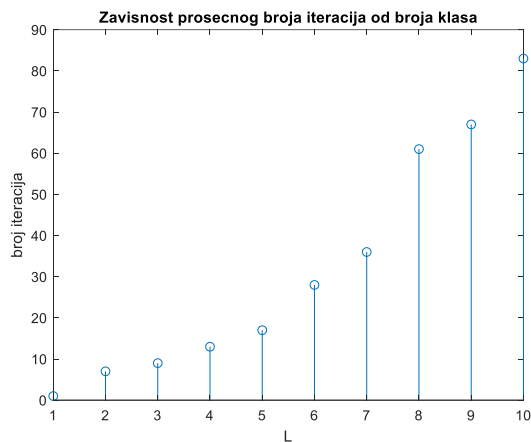
Slika 4.a.2: Inicijalna klasterizacija je stohastička. Ovo je prihvatljiva inicijalizacija za izabranu cmean klasterizaciju, a u to se možemo uveriti i na slici 4.3.



Slika 4.a.3: Kao što je već pomenuto, kvalitet cmean klasifikacije ne zavisi od početne klasterizacije, što je i eksperimentalno dokazano.

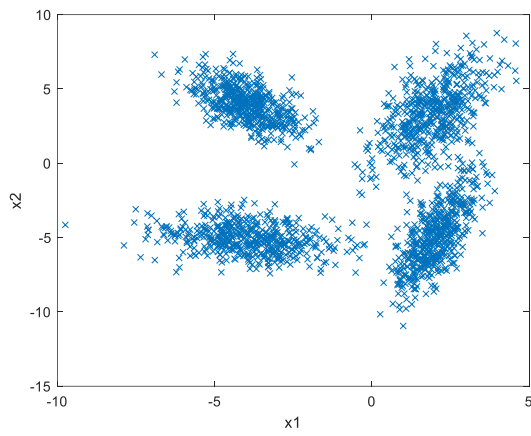


Slika 4.a.4: Za razliku od početne klasterizacije, osetljivost zavisi od maksimalnog broja iteracija. Za manje od oko 16 iteracija, algoritam ne uspeva da dođe do zadovoljavajućeg rešenja, međutim, za sve veće brojeve iteracija, rezultati su zadovoljavajući. Prosečan broj iteracija koji je bio potreban ovom algoritmu do konvergencije za četiri klase je 16.

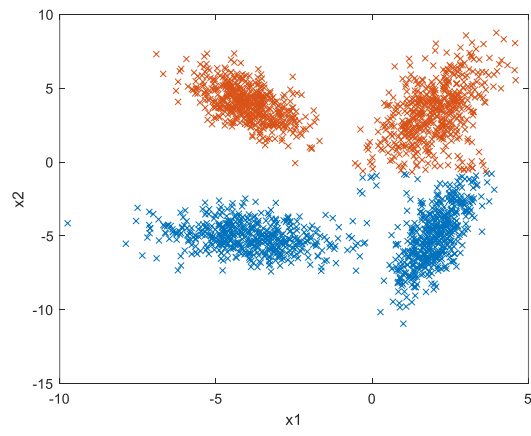


Slika 4.a.5: Vidimo da što je veći broj klasa, algoritmu je potrebno više iteracija da iskonvergira, što ima smisla jer se od njega traži da nađe nepostojeće razlike u zadatim odbircima – da nađe više klasa od stvarnog broja klasa.

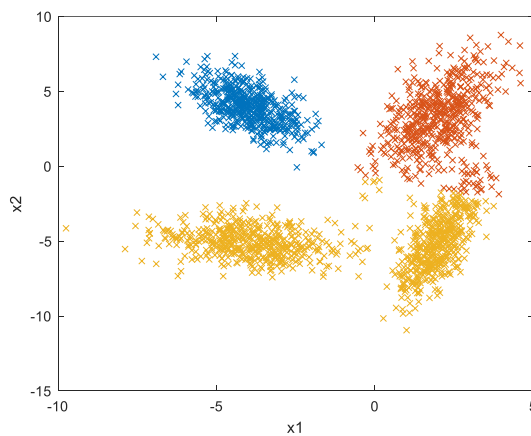
U nastavku su rezultati klasterizacije za različiti broj zadatih klasa (L). Različitim bojama obojeni su odbirci različitih klasa.



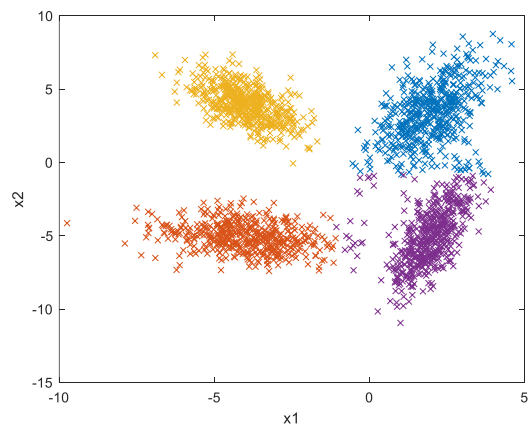
$L = 1$



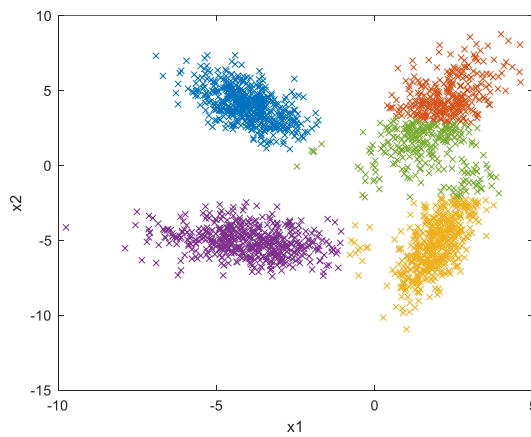
$L = 2$



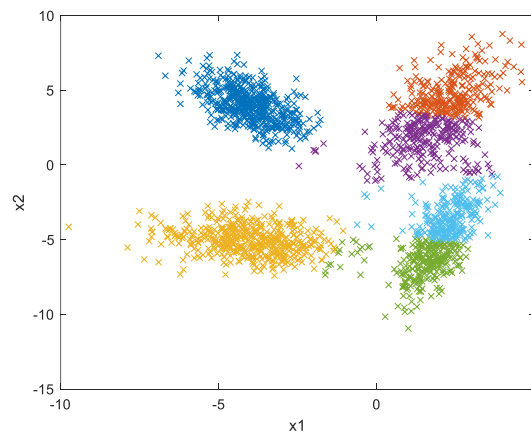
$L = 3$



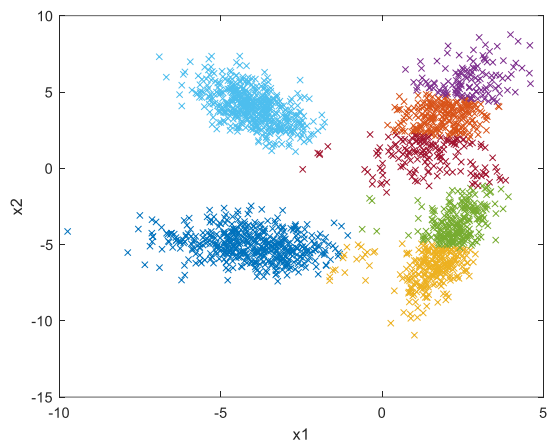
$L = 4$



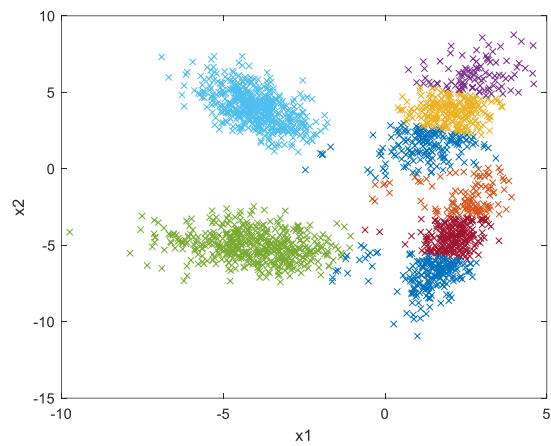
$L = 5$



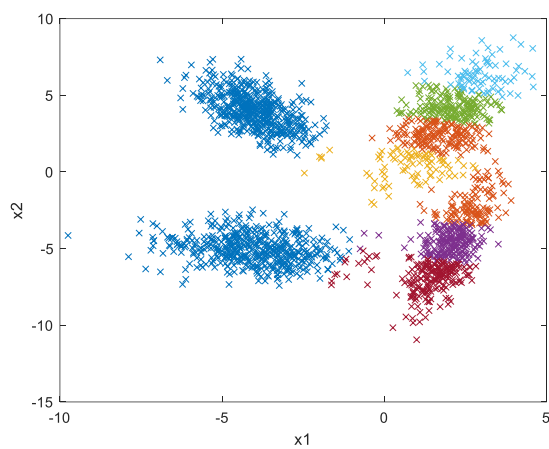
$L = 6$



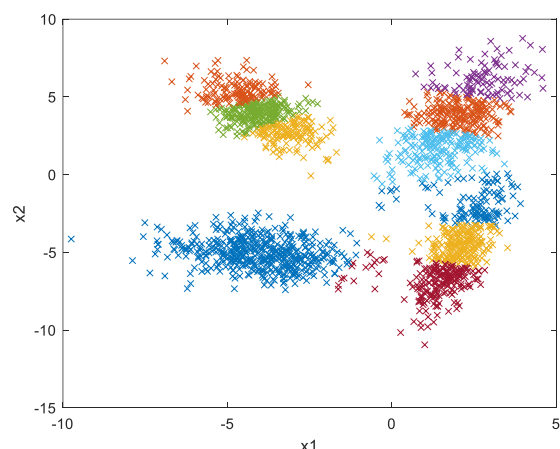
L = 7



L = 8

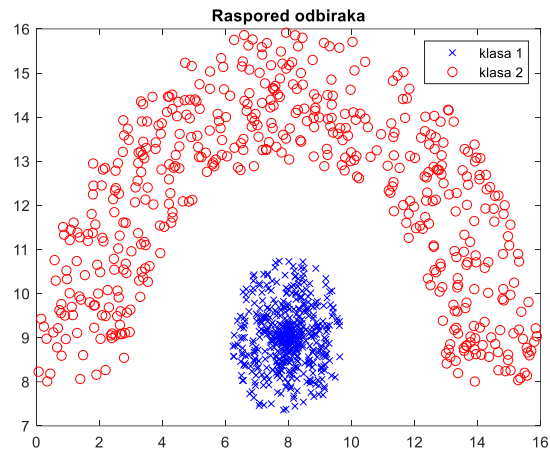


L = 9

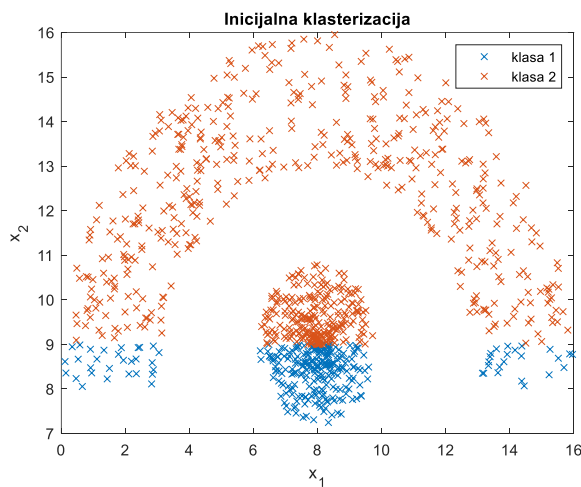


L = 10

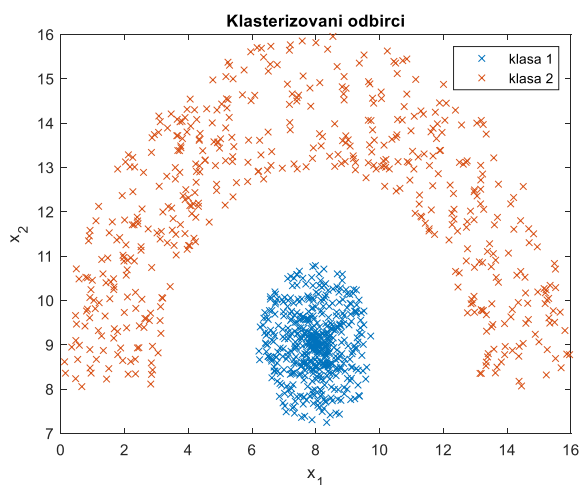
b)



Slika 4.b.1: Na slici je prikazan raspored odbiraka dve nelinearno separabilne klase. Ovog puta korišćićemo metodu normalne dekompozicije.



Slika 4.b.2: Za ovu metodu klasterizacije, inicijalna klasterizacija ne može biti sasvim slučajno odabrana. Ovde je ručno zadata početna klasterizacija, tako da prvoj klasi pripadaju svi odbirci sa drugom koordinatom manjom od 9, dok ostali odbirci pripadaju drugoj klasi. Mogla je biti upotrebljena i neka druga sofisticiranija metoda za odabir početne klasterizacije, međutim ova je dala zadovoljavajuće rezultate, pa je zadržana zbog jednostavnosti.



Slika 4.b.3: Klasterizacija je uspešno izvršena, sa osetljivošću 1 – svi odbirci su dobro klasifikovani. Broj potrebnih iteracija je 11.