Modeli zasnovani na širokom pojasu

Mašinsko učenje 2020/21.
Matematički fakultet

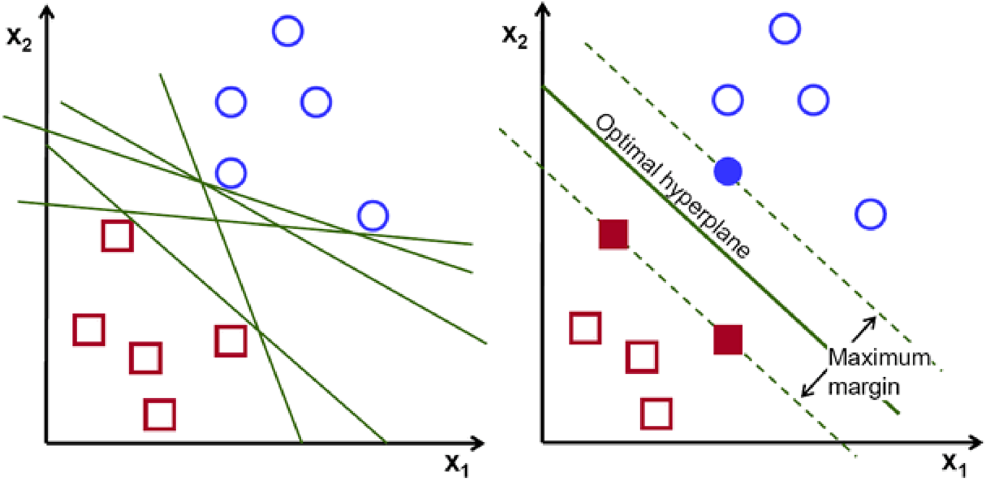
Pregled

Metod potpornih vektora za klasifikaciju

Metod potpornih vektora za regresiju

Algoritam k najbližih suseda zasnovan na širokom pojasu

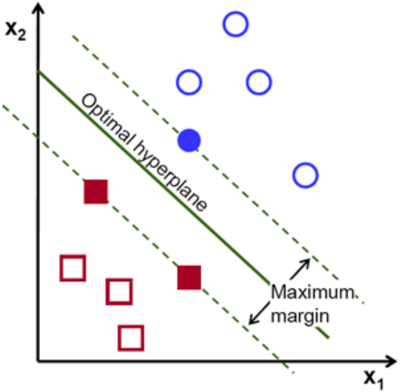
Metod potpornih vektora za klasifikaciju



Slika: OpenCV, Introduction to Support Vector Machines.

Metoda potpornih vektora (SVM)

- Razmatramo problem binarne klasifikacije u kom su klase označene sa -1 i 1

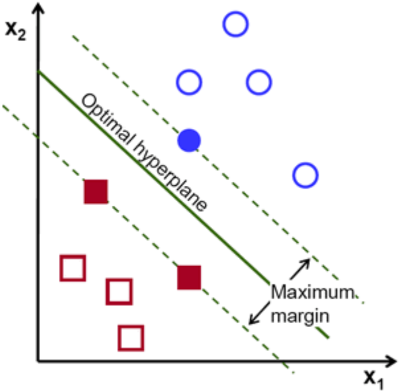


Metoda potpornih vektora (SVM)

- Razmatramo problem binarne klasifikacije u kom su klase označene sa -1 i 1
- Jednačina hiperravni:

$$w \cdot x + w_0 = 0$$

gde je w_0 slobodan član

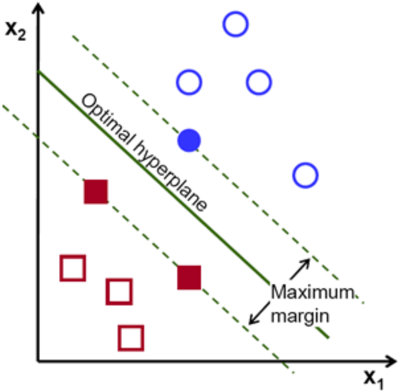


Metoda potpornih vektora (SVM)

- Razmatramo problem binarne klasifikacije u kom su klase označene sa -1 i 1
- Jednačina hiperravni:

$$w \cdot x + w_0 = 0$$

gde je w_0 slobodan član
- Među svim razdvajajućim hiperavnima, potrebno je naći optimalnu

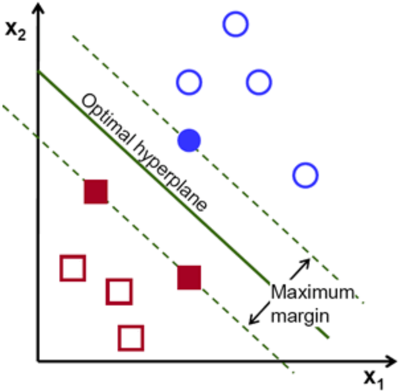


Metoda potpornih vektora (SVM)

- Razmatramo problem binarne klasifikacije u kom su klase označene sa -1 i 1
- Jednačina hiperravni:

$$w \cdot x + w_0 = 0$$

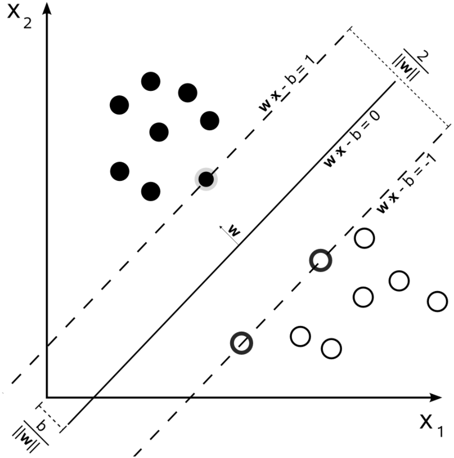
  gde je w_0 slobodan član
- Među svim razdvajajućim hiperavnima, potrebno je naći optimalnu
- *Optimalna hiperravan*, odnosno *hiperravan najšireg pojasa* je podjednako udaljena od najbližih predstavnika obe klase.



Ilustracija optimalne razdvajajuće hiperravni

- Optimalna hiperravan:

$$w \cdot x + w_0 = 0$$



Ilustracija optimalne razdvajajuće hiperravni

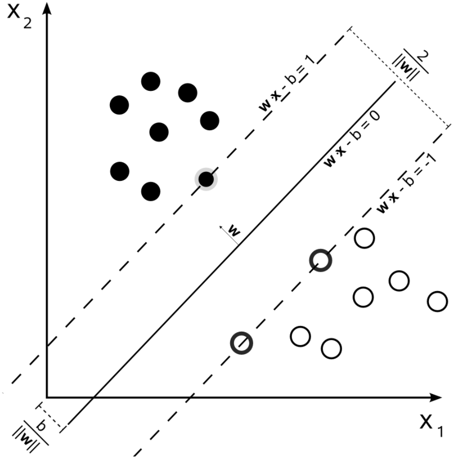
- Optimalna hiperravan:

$$w \cdot x + w_0 = 0$$

- Hiperravni paralelne optimalnoj:

$$w \cdot x + w_0 = c$$

$$w \cdot x + w_0 = -c$$



Ilustracija optimalne razdvajajuće hiperravni

- Optimalna hiperravan:

$$w \cdot x + w_0 = 0$$

- Hiperravni paralelne optimalnoj:

$$w \cdot x + w_0 = c$$

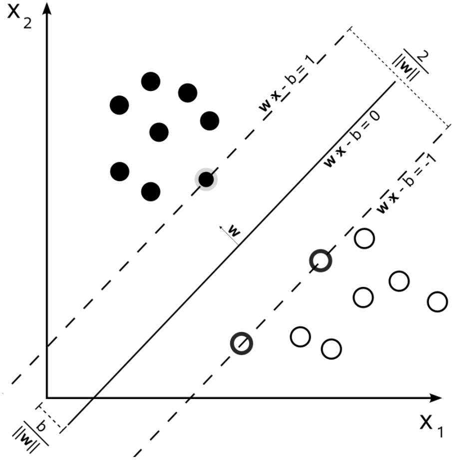
$$w \cdot x + w_0 = -c$$

- Podelimo sve sa c:

$$w \cdot x + w_0 = 0$$

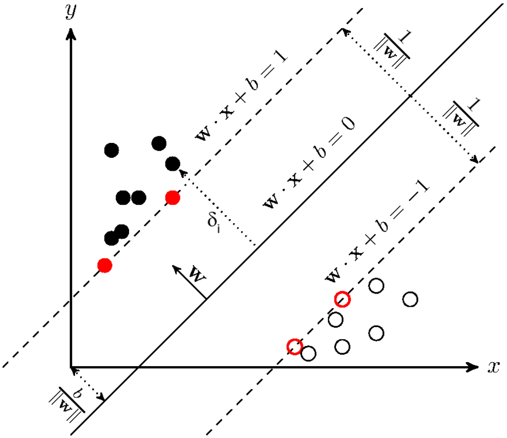
$$w \cdot x + w_0 = 1$$

$$w \cdot x + w_0 = -1$$



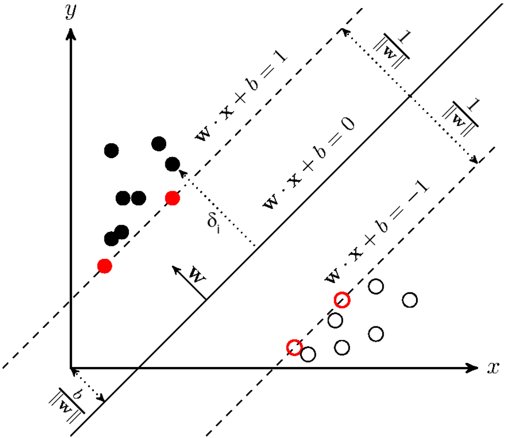
Ilustracija optimalne razdvajajuće hiperravni

- Rastojanje između optimalne hiperravni i jedne od pomenutih hiperravni koje su joj paralelne je upravo *pojas* koji treba da bude što veći



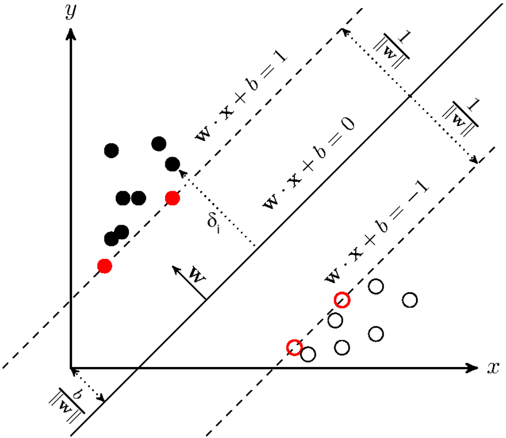
Ilustracija optimalne razdvajajuće hiperravni

- Rastojanje između optimalne hiperravni i jedne od pomenutih hiperravni koje su joj paralelne je upravo *pojas* koji treba da bude što veći
- Tačke koje se nalaze na pomenutim hiperravnima nazivaju se *potpornim vektorima*



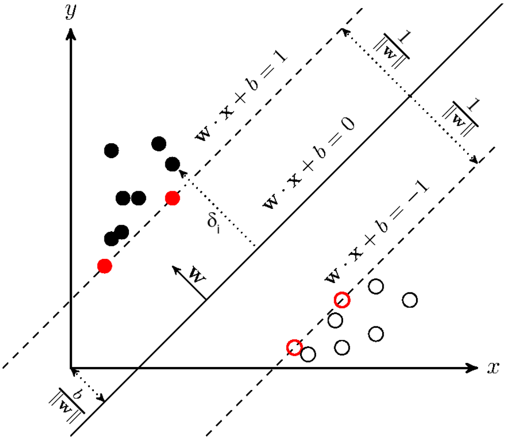
Ilustracija optimalne razdvajajuće hiperravni

- Rastojanje između optimalne hiperravni i jedne od pomenutih hiperravni koje su joj paralelne je upravo *pojas* koji treba da bude što veći
- Tačke koje se nalaze na pomenutim hiperravnima nazivaju se *potpornim vektorima*
- Za potporne vektore važi $|w \cdot x + w_0| = 1$



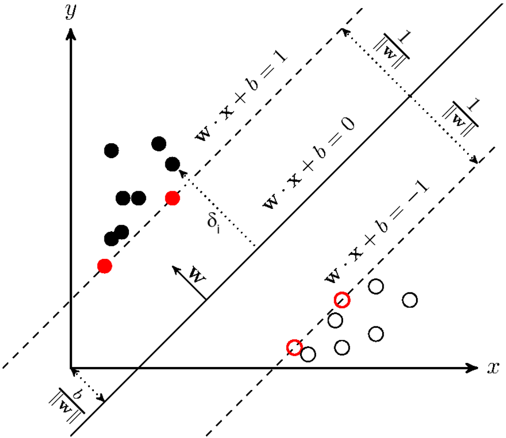
Ilustracija optimalne razdvajajuće hiperravni

- Za potporne vektore važi $|w \cdot x + w_0| = 1$



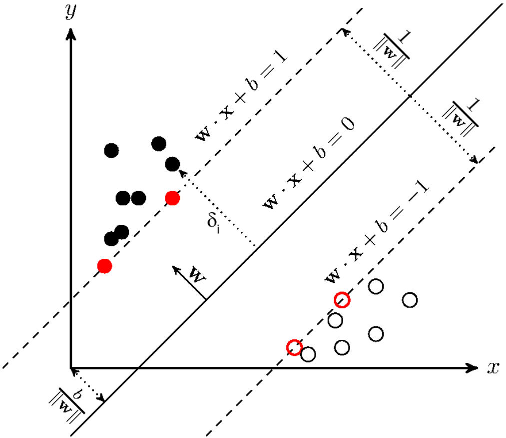
Ilustracija optimalne razdvajajuće hiperravni

- Za potporne vektore važi $|w \cdot x + w_0| = 1$
- Za ostale instance važi $w \cdot x + w_0 > 1$ ili $w \cdot x + w_0 < -1$



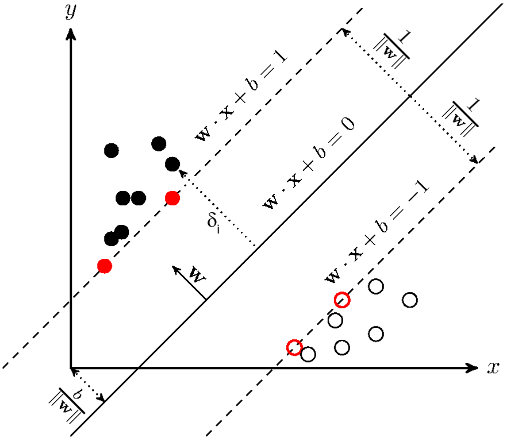
Ilustracija optimalne razdvajajuće hiperravni

- Za potporne vektore važi $|w \cdot x + w_0| = 1$
- Za ostale instance važi $w \cdot x + w_0 > 1$ ili $w \cdot x + w_0 < -1$
- Prilikom klasifikacije instance sa vektorom atributa x, klasa se određuje kao $sgn(w \cdot x + w_0)$



Ilustracija optimalne razdvajajuće hiperravni

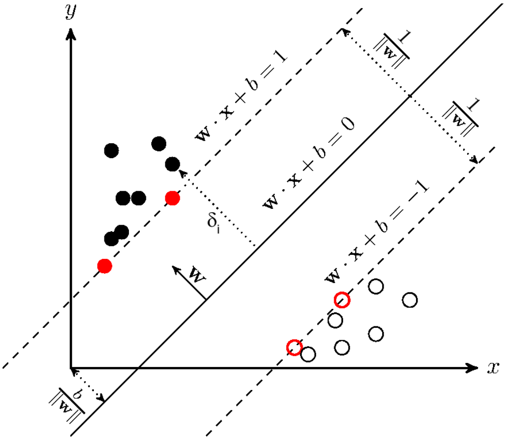
- Potrebno je da odredimo širinu pojasa koju želimo da maksimizujemo



Ilustracija optimalne razdvajajuće hiperravni

- Potrebno je da odredimo širinu pojasa koju želimo da maksimizujemo
- Računamo rastojanje između jednog potpornog vektora i optimalne hiperravni

$$\frac{|w \cdot x + w_0|}{\|w\|_2} = \frac{1}{\|w\|_2}$$



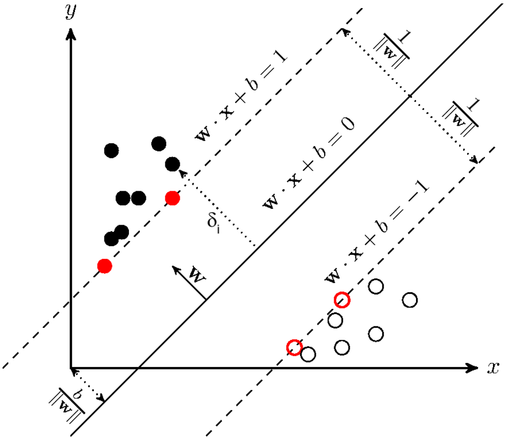
Ilustracija optimalne razdvajajuće hiperravni

- Potrebno je da odredimo širinu pojasa koju želimo da maksimizujemo
- Računamo rastojanje između jednog potpornog vektora i optimalne hiperravni

$$\frac{|w \cdot x + w_0|}{\|w\|_2} = \frac{1}{\|w\|_2}$$

- Širina pojasa:

$$\frac{2}{\|w\|_2}$$



Osnovna formulacija

- Ukupno rastojanje između klasa, u pravcu normalnom u odnosu na optimalnu hiperravan je $2/\|w\|$

Osnovna formulacija

- Ukupno rastojanje između klasa, u pravcu normalnom u odnosu na optimalnu hiperravan je $2/\|w\|$
- Optimalna hiperravan se dobija pronalaženjem koeficijenata koji maksimizuju ovaj izraz pod uslovima da su sve tačke sa pravih strana te hiperravni

Osnovna formulacija

- Ukupno rastojanje između klasa, u pravcu normalnom u odnosu na optimalnu hiperravan je $2/\|w\|$
- Optimalna hiperravan se dobija pronalaženjem koeficijenata koji maksimizuju ovaj izraz pod uslovima da su sve tačke sa pravih strana te hiperravni
- Optimizacioni problem:

$$\min_{w, w_0} \frac{\|w\|}{2}$$

$$y_i \left(w \cdot x_i + w_0 \right) \geq 1 \qquad i = 1, \ldots, N$$

Osnovna formulacija

- Ukupno rastojanje između klasa, u pravcu normalnom u odnosu na optimalnu hiperravan je $2/\|w\|$
- Optimalna hiperravan se dobija pronalaženjem koeficijenata koji maksimizuju ovaj izraz pod uslovima da su sve tačke sa pravih strana te hiperravni
- Optimizacioni problem:

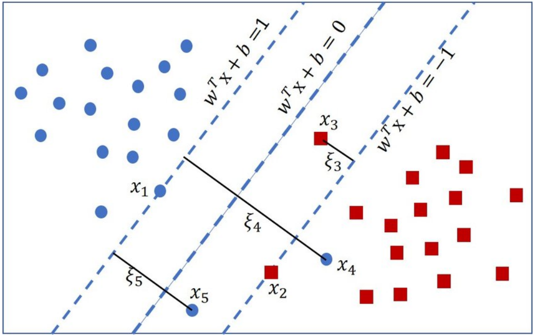
$$\min_{w,w_0} \frac{\|w\|}{2}$$

$$y_i\left(w \cdot x_i + w_0\right) \geq 1 \qquad i = 1, \ldots, N$$

- Dodatni uslovi izražavaju potrebu da sve tačke budu na većem rastojanju od optimalne hiperravni nego što su potporni vektori koji su na rastojanju 1

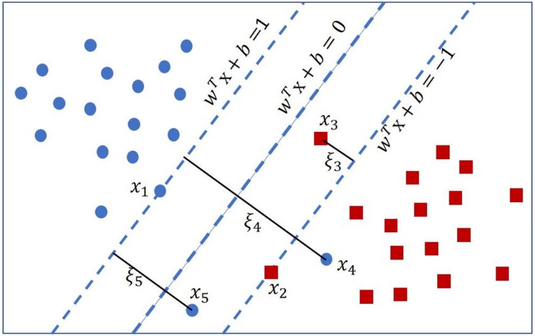
Linearno neseparabilan slučaj

- U praksi retko možemo očekivati linearnu razdvojivost klasa



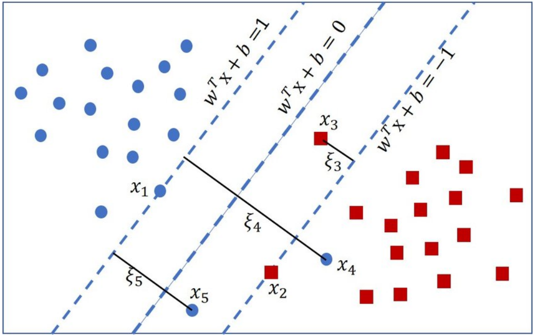
Linearno neseparabilan slučaj

- U praksi retko možemo očekivati linearnu razdvojivost klasa
- Otud je neophodno prihvatiti neke greške, uz zahtev da budu što manje



Linearno neseparabilan slučaj

- U praksi retko možemo očekivati linearnu razdvojivost klasa
- Otud je neophodno prihvatiti neke greške, uz zahtev da budu što manje
- Uvodimo nove promenljive ξ_i za svaku instancu u skupu za obučavanje, koje mere koliko je svaka instanca daleko od hiperravni određene potpornim vektorima njene klase, ali samo pod pretpostavkom da je sa pogrešne strane

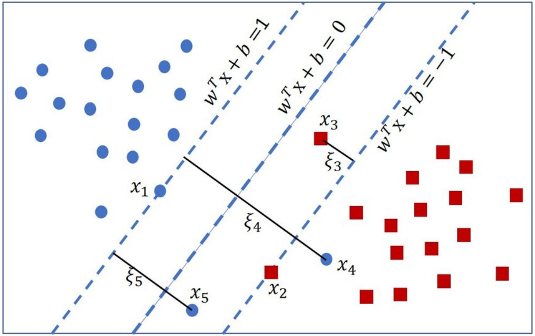


Metod potpornih vektora sa mekim pojasom

$$\min_{w,w_0} \frac{\|w\|_2}{2} + C\left(\sum_{i=1}^{N} \xi_i\right)$$

$$y_i\left(w \cdot x_i + w_0\right) \geq 1 - \xi_i, \qquad i = 1, \ldots, N$$

$$\xi_i \geq 0, \qquad i = 1, \ldots, N$$



Gde je funkcija greške, a gde regularizacija?

$$\min_{w,w_0} \frac{\|w\|_2}{2} + C\left(\sum_{i=1}^{N} \xi_i\right)$$

$$y_i\left(w \cdot x_i + w_0\right) \geq 1 - \xi_i, \qquad i = 1, \ldots, N$$

$$\xi_i \geq 0, \qquad i = 1, \ldots, N$$

- Možemo pokazati da važi:

$$\xi_i = \max(0, 1 - y_i(w \cdot x_i + w_0))$$

Gde je funkcija greške, a gde regularizacija?

$$\min_{w,w_0} \frac{\|w\|_2}{2} + C\left(\sum_{i=1}^{N} \xi_i\right)$$

$$y_i\left(w \cdot x_i + w_0\right) \geq 1 - \xi_i, \qquad i = 1, \ldots, N$$

$$\xi_i \geq 0, \qquad i = 1, \ldots, N$$

- Možemo pokazati da važi:

$$\xi_i = \max(0, 1 - y_i(w \cdot x_i + w_0))$$

- Tada se optimizacioni problem može zapisati:

$$\min_{w,w_0} \frac{\|w\|_2}{2} + C\sum_{i=1}^{N} \max(0, 1 - y_i(w \cdot x_i + w_0))$$

Gde je funkcija greške, a gde regularizacija?

$$\min_{w,w_0} \frac{\|w\|_2}{2} + C\left(\sum_{i=1}^{N} \xi_i\right)$$

$$y_i\left(w \cdot x_i + w_0\right) \geq 1 - \xi_i, \qquad i = 1, \ldots, N$$

$$\xi_i \geq 0, \qquad i = 1, \ldots, N$$

- Možemo pokazati da važi:

$$\xi_i = \max(0, 1 - y_i(w \cdot x_i + w_0))$$

- Tada se optimizacioni problem može zapisati:

$$\min_{w,w_0} \frac{\|w\|_2}{2} + C\sum_{i=1}^{N} \max(0, 1 - y_i(w \cdot x_i + w_0))$$

- Što je ekvivalentno sa:

$$\min_{w,w_0} \sum_{i=1}^{N} \max(0, 1 - y_i(w \cdot x_i + w_0)) + \lambda\|w\|_2$$

Funkcija greške u vidu šarke

$$\min_{w,w_0} \sum_{i=1}^{N} \max(0, 1 - y_i(w \cdot x_i + w_0)) + \lambda \|w\|_2$$

- u je stvarna vrednost (može biti 1 ili -1), v predviđanje (neprekidna vrednost $w \cdot x_i + w_0$), posmatrajmo funkciju

$$L(u, v) = \max(0, 1 - uv)$$

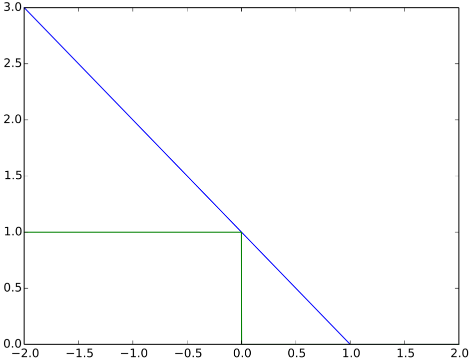
- Da li je ova funkcija smislena funkcija greške?

Funkcija greške u vidu šarke i indikatorska funkcija

u je stvarna vrednost (može biti 1 ili -1), v predviđanje (neprekidna vrednost $w \cdot x_i + w_0$)

plavo: $L_{hinge}(u, v) = \max(0, 1 - uv)$

zeleno: $L_{ind}(u, v) = I(u \neq v)$



Rešenje optimizacionog problema

- Može se pokazati da je rešenje optimizacionog problema oblika:

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

gde su α_i Lagranžovi množioci za koje važi $0 \leq \alpha_i \leq C$

- Podaci x_i za koje je $\alpha_i > 0$ su *potporni vektori*
- Model je onda oblika

$$f_{w,w_0}(x) = \sum_{i=1}^{N} \alpha_i y_i \ x_i \cdot x + w_0$$

- Klasa se određuje funkcijom $\mathrm{sgn}(f_{w,w_0}(x))$

Poređenje metode potpornih vektora i logističke regresije

- Da bismo uporedili dve metode, pratimo shemu dizajna modela mašinskog učenja

Poređenje metode potpornih vektora i logističke regresije

- Da bismo uporedili dve metode, pratimo shemu dizajna modela mašinskog učenja
- Vrsta modela - diskriminativni modeli klasifikacije, jedan je probabilistički, drugi nije

Poređenje metode potpornih vektora i logističke regresije

- Da bismo uporedili dve metode, pratimo shemu dizajna modela mašinskog učenja
- Vrsta modela - diskriminativni modeli klasifikacije, jedan je probabilistički, drugi nije
- Forma modela - hiperravan u oba slučaja

Poređenje metode potpornih vektora i logističke regresije

- Da bismo uporedili dve metode, pratimo shemu dizajna modela mašinskog učenja
- Vrsta modela - diskriminativni modeli klasifikacije, jedan je probabilistički, drugi nije
- Forma modela - hiperravan u oba slučaja
- Funkcija greške - različite

Poređenje metode potpornih vektora i logističke regresije

- Da bismo uporedili dve metode, pratimo shemu dizajna modela mašinskog učenja
- Vrsta modela - diskriminativni modeli klasifikacije, jedan je probabilistički, drugi nije
- Forma modela - hiperravan u oba slučaja
- Funkcija greške - različite
- Regularizacija - kod metode potpornih vektora je ugrađena po konstrukciji a kod logističke regresije nije; ipak, ovo nije suštinska razlika jer se regularizacija uvek može dodati

Poređenje metode potpornih vektora i logističke regresije

- Da bismo uporedili dve metode, pratimo shemu dizajna modela mašinskog učenja
- Vrsta modela - diskriminativni modeli klasifikacije, jedan je probabilistički, drugi nije
- Forma modela - hiperravan u oba slučaja
- Funkcija greške - različite
- Regularizacija - kod metode potpornih vektora je ugrađena po konstrukciji a kod logističke regresije nije; ipak, ovo nije suštinska razlika jer se regularizacija uvek može dodati
- Optimizacioni model - bitan za brzinu obučavanja ali ukoliko je optimizacija uspešno završena približnim nalaženjem minimuma, nema posledica po ponašanje modela pri predviđanju

Poređenje metode potpornih vektora i logističke regresije

- Da bismo uporedili dve metode, pratimo shemu dizajna modela mašinskog učenja
- Vrsta modela - diskriminativni modeli klasifikacije, jedan je probabilistički, drugi nije
- Forma modela - hiperravan u oba slučaja
- Funkcija greške - različite
- Regularizacija - kod metode potpornih vektora je ugrađena po konstrukciji a kod logističke regresije nije; ipak, ovo nije suštinska razlika jer se regularizacija uvek može dodati
- Optimizacioni model - bitan za brzinu obučavanja ali ukoliko je optimizacija uspešno završena približnim nalaženjem minimuma, nema posledica po ponašanje modela pri predviđanju
- Zaključak - razlika potiče od probabilističke prirode logističke regresije ili od funkcije greške

Poređenje metode potpornih vektora i logističke regresije

- Kako ta probabilistička priroda utiče na izbor modela?

Poređenje metode potpornih vektora i logističke regresije

- Kako ta probabilistička priroda utiče na izbor modela?
- Konkretna probabilistička pretpostavka je vezana za Bernulijevu raspodelu i metod maksimalne verodostojnosti.

Poređenje metode potpornih vektora i logističke regresije

- Kako ta probabilistička priroda utiče na izbor modela?
- Konkretna probabilistička pretpostavka je vezana za Bernulijevu raspodelu i metod maksimalne verodostojnosti.
- Pažljivom analizom izvođenja optimizacionog problema, može se videti da ona direktno uslovljava izbor funkcije greške.

Poređenje metode potpornih vektora i logističke regresije

- Kako ta probabilistička priroda utiče na izbor modela?
- Konkretna probabilistička pretpostavka je vezana za Bernulijevu raspodelu i metod maksimalne verodostojnosti.
- Pažljivom analizom izvođenja optimizacionog problema, može se videti da ona direktno uslovljava izbor funkcije greške.
- Zaključujemo da se obe razlike svode na razliku u funkciji greške!

Poređenje metode potpornih vektora i logističke regresije

- Kako bi takva analiza bila moguća, potrebno je eliminisati određene razlike u formulacijama.

Poređenje metode potpornih vektora i logističke regresije

- Kako bi takva analiza bila moguća, potrebno je eliminisati određene razlike u formulacijama.
- Metod potpornih vektora podrazumeva da su oznake klasa 1 i -1, dok logistička regresija podrazumeva oznake 0 i 1.

Poređenje metode potpornih vektora i logističke regresije

- Kako bi takva analiza bila moguća, potrebno je eliminisati određene razlike u formulacijama.
- Metod potpornih vektora podrazumeva da su oznake klasa 1 i -1, dok logistička regresija podrazumeva oznake 0 i 1.
- Može se pokazati da u slučaju oznaka 1 i -1 problem logističke regresije ima formu

$$\min_w \sum_{i=1}^{N} \log(1 + e^{-y_i w \cdot x_i})$$

odnosno da je funkcija greške

$$L_{LR}(u, v) = \log(1 + e^{-uv})$$

a u slučaju metoda potpornih vektora, to je

$$L_{SV}(u, v) = \max(0, 1 - uv)$$

Poređenje metode potpornih vektora i logističke regresije

- Kako bi se lakše uporedile, podelimo funkciju greške logističke regresije sa $\log 2$, tako da obe prolaze kroz tačku $(0, 1)$ (množenje konstantom pri minimizaciji nebitno)

Poređenje metode potpornih vektora i logističke regresije

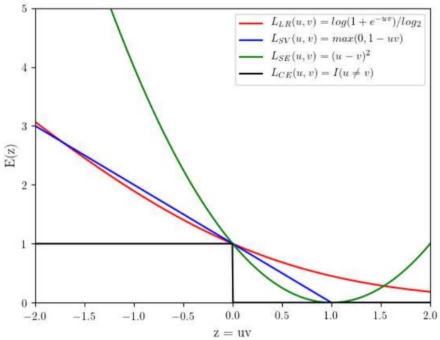
- Kako bi se lakše uporedile, podelimo funkciju greške logističke regresije sa $\log 2$, tako da obe prolaze kroz tačku $(0, 1)$ (množenje konstantom pri minimizaciji nebitno)
- Poređenja radi, dodajmo u poređenje i metod koji bi radio na osnovu kvadratne funkcije greške
$$L_{SE}(u, v) = (u - v)^2$$
pošto ju je zaista moguće primeniti i u slučaju binarne klasifikacije kada su oznake klasa numeričke

Poređenje metode potpornih vektora i logističke regresije

- Kako bi se lakše uporedile, podelimo funkciju greške logističke regresije sa $\log 2$, tako da obe prolaze kroz tačku $(0, 1)$ (množenje konstantom pri minimizaciji nebitno)
- Poređenja radi, dodajmo u poređenje i metod koji bi radio na osnovu kvadratne funkcije greške
$$L_{SE}(u, v) = (u - v)^2$$
pošto ju je zaista moguće primeniti i u slučaju binarne klasifikacije kada su oznake klasa numeričke
- Uzmimo u obzir i standardnu grešku klasifikacije
$$L_{CE}(u, v) = I(u \neq v)$$

Poređenje metode potpornih vektora i logističke regresije

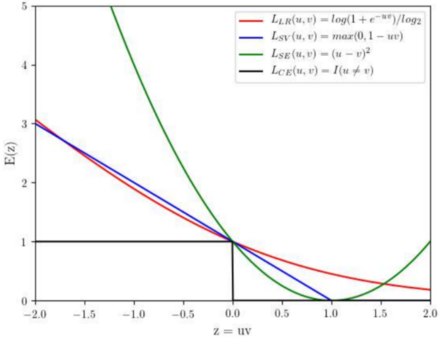
- Na horizontalnoj osi je prikazana vrednost poizvoda uv, a na vertikalnoj odgovarajuća vrednost greške.



Slika 4.4: Grafici četiri funkcije greške – logističke (crveno), metoda potpornih vektora (plavo), kvadratne (zeleno) i greške klasifikacije (crno).

Poređenje metode potpornih vektora i logističke regresije

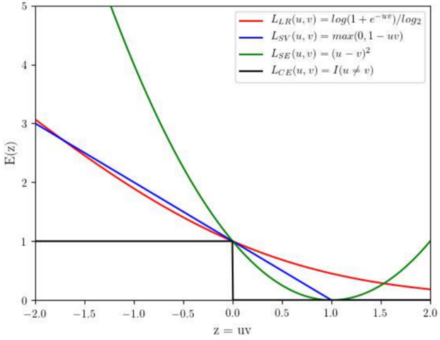
- Na horizontalnoj osi je prikazana vrednost poizvoda uv, a na vertikalnoj odgovarajuća vrednost greške.
- Pozitivne vrednosti uv odgovaraju tačno klasifikovanim instancama, a negativne pogrešno klasifikovanim.



Slika 4.4: Grafici četiri funkcije greške – logističke (crveno), metoda potpornih vektora (plavo), kvadratne (zeleno) i greške klasifikacije (crno).

Poređenje metode potpornih vektora i logističke regresije

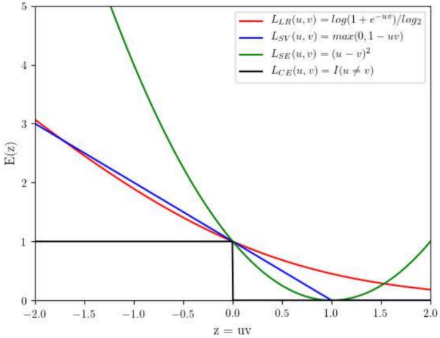
- Očigledno, logistička funkcija greške kažnjava ne samo pogrešno, već i ispravno klasifikovane instance, čak i ako su daleko od razdvajajuće hiperravni



Slika 4.4: Grafici četiri funkcije greške – logističke (crveno), metoda potpornih vektora (plavo), kvadratne (zeleno) i greške klasifikacije (crno).

Poređenje metode potpornih vektora i logističke regresije

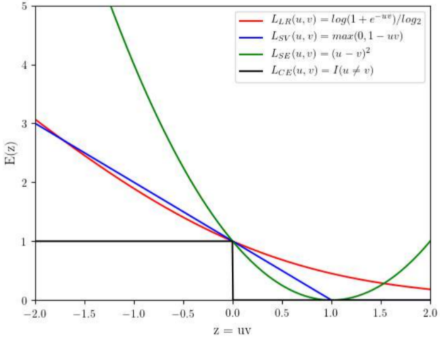
- Očigledno, logistička funkcija greške kažnjava ne samo pogrešno, već i ispravno klasifikovane instance, čak i ako su daleko od razdvajajuće hiperravni
- To je zbog toga što funkcija greške za logističku regresiju nikada nije 0 već beskonačno teži nuli



Slika 4.4: Grafici četiri funkcije greške – logističke (crveno), metoda potpornih vektora (plavo), kvadratne (zeleno) i greške klasifikacije (crno).

Poređenje metode potpornih vektora i logističke regresije

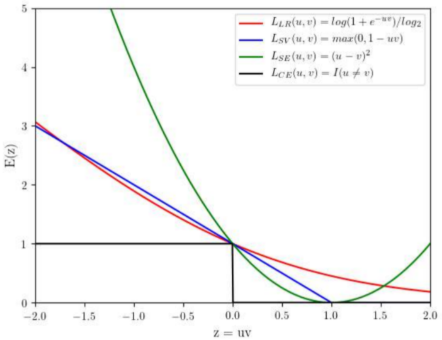
- Očigledno, logistička funkcija greške kažnjava ne samo pogrešno, već i ispravno klasifikovane instance, čak i ako su daleko od razdvajajuće hiperravni
- To je zbog toga što funkcija greške za logističku regresiju nikada nije 0 već beskonačno teži nuli
- To vodi tome da se kod logističke regresije ne bira hiperravan najšireg pojasa, kao i da sve instance učestvuju u definisanju modela.



Slika 4.4: Grafici četiri funkcije greške – logističke (crveno), metoda potpornih vektora (plavo), kvadratne (zeleno) i greške klasifikacije (crno).

Poređenje metode potpornih vektora i logističke regresije

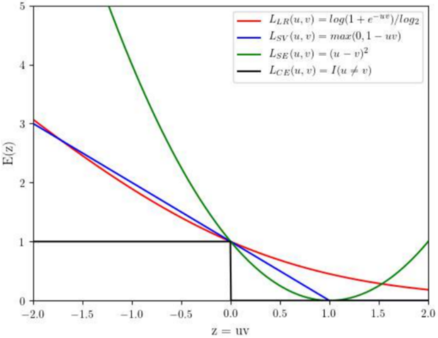
- Funkcija greške metoda potpornih vektora dostiže nulu



Slika 4.4: Grafici četiri funkcije greške – logističke (crveno), metoda potpornih vektora (plavo), kvadratne (zeleno) i greške klasifikacije (crno).

Poređenje metode potpornih vektora i logističke regresije

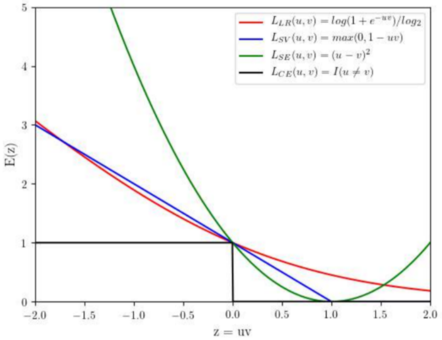
- Funkcija greške metoda potpornih vektora dostiže nulu
- Upravo to vodi kako maksimalnom odstojanju, tako i zanemarivanju većine podataka (značajni su samo potporni vektori)



Slika 4.4: Grafici četiri funkcije greške – logističke (crveno), metoda potpornih vektora (plavo), kvadratne (zeleno) i greške klasifikacije (crno).

Poređenje metode potpornih vektora i logističke regresije

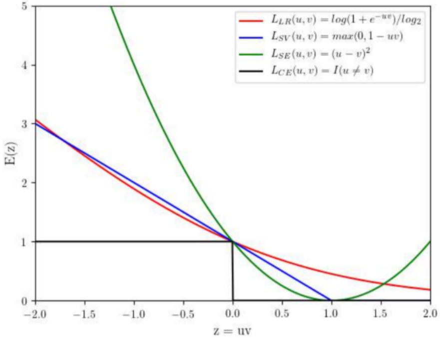
- Još jedna razlika je da logistička funkcija greške raste brže što je *uv* manje, pa je stoga i osetljivija na odudarajuće podatke.



Slika 4.4: Grafici četiri funkcije greške – logističke (crveno), metoda potpornih vektora (plavo), kvadratne (zeleno) i greške klasifikacije (crno).

Poređenje metode potpornih vektora i logističke regresije

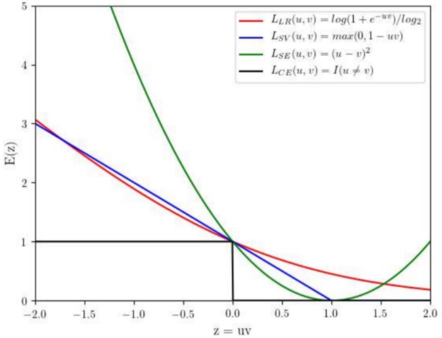
- Sve ove funkcije predstavljaju nekakvu aproksimaciju (indikatorske) funkcije greške klasifikacije.



Slika 4.4: Grafici četiri funkcije greške – logističke (crveno), metoda potpornih vektora (plavo), kvadratne (zeleno) i greške klasifikacije (crno).

Poređenje metode potpornih vektora i logističke regresije

- Sve ove funkcije predstavljaju nekakvu aproksimaciju (indikatorske) funkcije greške klasifikacije.
- Kvadratna funkcija greške očito predstavlja vrlo lošu aproksimaciju, pošto jako kažnjava i ispravno klasifikovane tačke, čim je proizvod *uv* veći od 1.



Slika 4.4: Grafici četiri funkcije greške – logističke (crveno), metoda potpornih vektora (plavo), kvadratne (zeleno) i greške klasifikacije (crno).

Pregled

Metod potpornih vektora za klasifikaciju

Metod potpornih vektora za regresiju

Algoritam k najbližih suseda zasnovan na širokom pojasu

Metod potpornih vektora za regresiju

- Potrebno je izloženi metod potpornih vektora za klasifikaciju prilagoditi za regresioni problem

Metod potpornih vektora za regresiju

- Potrebno je izloženi metod potpornih vektora za klasifikaciju prilagoditi za regresioni problem
- Imamo tačke i želimo da konstruišemo hiperravan koja prolazi kroz sve te tačke

Metod potpornih vektora za regresiju

- Potrebno je izloženi metod potpornih vektora za klasifikaciju prilagoditi za regresioni problem
- Imamo tačke i želimo da konstruišemo hiperravan koja prolazi kroz sve te tačke
- Jedan način za prilagođavanje bi bio sledeći:

$$\min_{w} \|w\|_2^2$$

$$|w \cdot x + w_0 - y| = 0$$

Metod potpornih vektora za regresiju

- Jedna važna tehnička razlika u odnosu na klasifikaciju je u tome što čak ni u osnovnoj varijanti metoda nema smisla tražiti tačna predviđanja, što je u linearno razdvojivom slučaju kod klasifikacije bio zahtev.

Metod potpornih vektora za regresiju

- Jedna važna tehnička razlika u odnosu na klasifikaciju je u tome što čak ni u osnovnoj varijanti metoda nema smisla tražiti tačna predviđanja, što je u linearno razdvojivom slučaju kod klasifikacije bio zahtev.
- Naime, kod binarne klasifikacije postoje dva moguća ishoda – 1 i -1 i sve vrednosti koje linearni model daje zaokružuju se na njih.

Metod potpornih vektora za regresiju

- Jedna važna tehnička razlika u odnosu na klasifikaciju je u tome što čak ni u osnovnoj varijanti metoda nema smisla tražiti tačna predviđanja, što je u linearno razdvojivom slučaju kod klasifikacije bio zahtev.
- Naime, kod binarne klasifikacije postoje dva moguća ishoda – 1 i -1 i sve vrednosti koje linearni model daje zaokružuju se na njih.
- U slučaju regresije nije potrebno da model da baš vrednost 1 ili -1 već postoji kontinuum ishoda i zahtev za tačnom jednakošću je prejak

Metod potpornih vektora za regresiju

- Jedna važna tehnička razlika u odnosu na klasifikaciju je u tome što čak ni u osnovnoj varijanti metoda nema smisla tražiti tačna predviđanja, što je u linearno razdvojivom slučaju kod klasifikacije bio zahtev.
- Naime, kod binarne klasifikacije postoje dva moguća ishoda – 1 i -1 i sve vrednosti koje linearni model daje zaokružuju se na njih.
- U slučaju regresije nije potrebno da model da baš vrednost 1 ili -1 već postoji kontinuum ishoda i zahtev za tačnom jednakošću je prejak
- Dodatno, često bi mogao biti i štetan: podaci retko predstavljaju merenja promenljivih veličina sa savršenom tačnošću; ako podaci sadrže grešku, nema smisla insistirati da se ta greška nauči

Metod potpornih vektora za regresiju

- Jedna važna tehnička razlika u odnosu na klasifikaciju je u tome što čak ni u osnovnoj varijanti metoda nema smisla tražiti tačna predviđanja, što je u linearno razdvojivom slučaju kod klasifikacije bio zahtev.
- Naime, kod binarne klasifikacije postoje dva moguća ishoda – 1 i -1 i sve vrednosti koje linearni model daje zaokružuju se na njih.
- U slučaju regresije nije potrebno da model da baš vrednost 1 ili -1 već postoji kontinuum ishoda i zahtev za tačnom jednakošću je prejak
- Dodatno, često bi mogao biti i štetan: podaci retko predstavljaju merenja promenljivih veličina sa savršenom tačnošću; ako podaci sadrže grešku, nema smisla insistirati da se ta greška nauči
- Stoga, uvodi se parametar tolerancije ε koji izražava razliku između predviđanja i stvarne vrednosti koja se smatra potpuno prihvatljivom

Metod potpornih vektora za regresiju

- Osnovni model izgleda ovako:

$$\min_{w} \|w\|_2^2$$

$$|w \cdot x_i + w_0 - y_i| \leq \varepsilon \qquad i = 1, \ldots, N$$

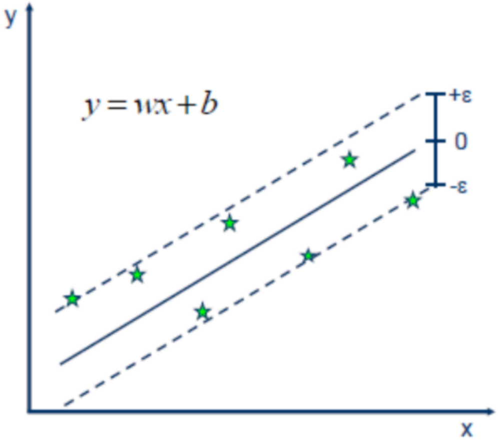
odnosno

$$\min_{w} \|w\|_2^2$$

$$w \cdot x_i + w_0 - y_i < \varepsilon \qquad i = 1, \ldots, N$$

$$y_i - w \cdot x_i - w_0 < \varepsilon \qquad i = 1, \ldots, N$$

Metod potpornih vektora za regresiju



Ograničenja nalažu da predviđanja ne mogu biti daleko od pravih vrednosti, dok minimizacija norme sprečava izbor modela koji brzo menja vrednosti, odnosno umanjuje prilagodljivost.

Metod potpornih vektora za regresiju

- Ova formulacija, kao i u slučaju klasifikacionog problema, ima nedostatak da ne dozvoljava greške u predviđanjima (osim za fiksiranu vrednost ε)

Metod potpornih vektora za regresiju

- Ova formulacija, kao i u slučaju klasifikacionog problema, ima nedostatak da ne dozvoljava greške u predviđanjima (osim za fiksiranu vrednost ε)
- Slično slučaju mekog pojasa, tolerancija na greške se omogućava uvođenjem novih promenljivih:

$$\min_{w} \|w\|_2^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$

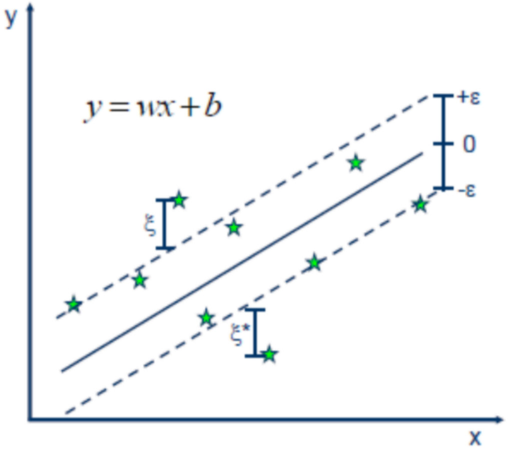
$$w \cdot x_i + w_0 - y_i < \varepsilon + \xi_i \qquad i = 1, \ldots, N$$

$$y_i - w \cdot x_i - w_0 < \varepsilon + \xi_i^* \qquad i = 1, \ldots, N$$

$$\xi_i \geq 0 \qquad i = 1, \ldots, N$$

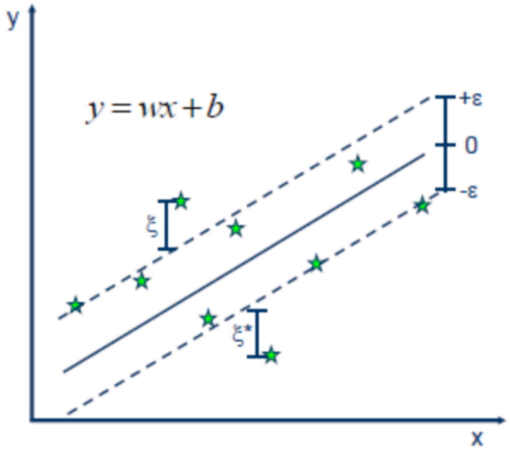
$$\xi_i^* \geq 0 \qquad i = 1, \ldots, N$$

Metod potpornih vektora za regresiju



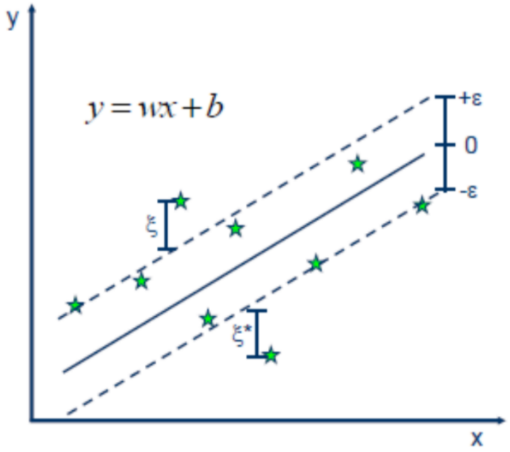
Široki pojas kod regresije

- Gde je u slučaju regresije široki pojas?



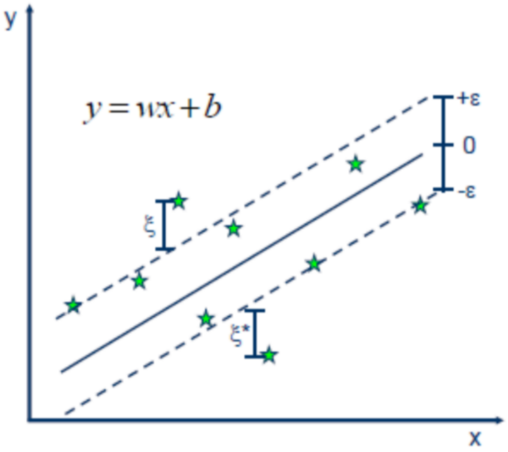
Široki pojas kod regresije

- Gde je u slučaju regresije široki pojas?
- Da li je možda u ε okolini modela?



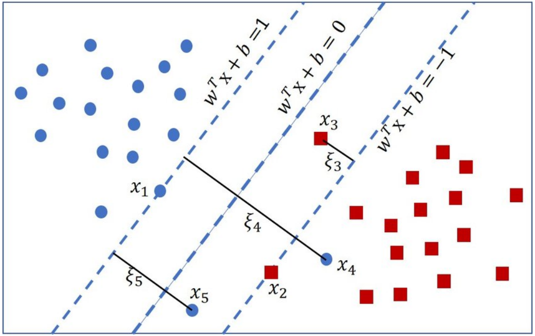
Široki pojas kod regresije

- Gde je u slučaju regresije široki pojas?
- Da li je možda u ε okolini modela?
- Ne, jer u idealnom slučaju baš u ε okolini se nalaze svi podaci, a smisao širokog pojasa je da u idealnom slučaju baš u njemu nema podataka.



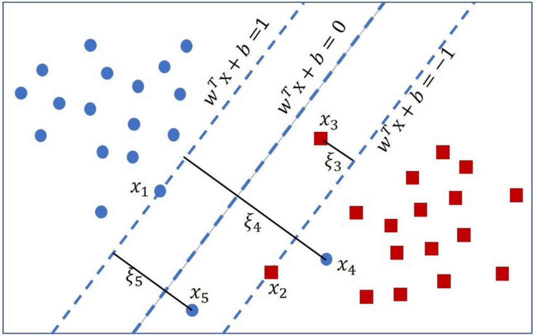
Široki pojas kod regresije

- Setimo se kako funkcija greške kažnjava instance u slučaju klasifikacije: potporni vektori i instance koje su dalje od njih u odgovarajućem poluprostoru ne doprinose grešci.



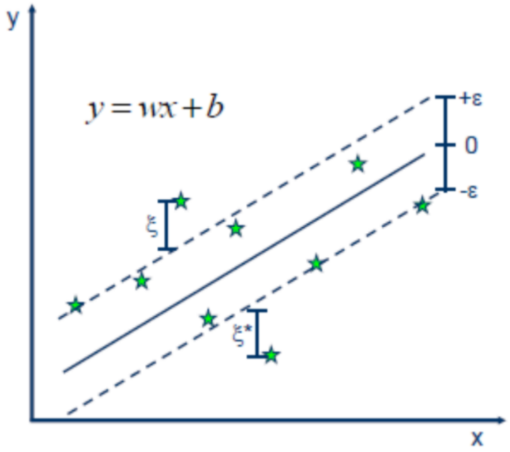
Široki pojas kod regresije

- Setimo se kako funkcija greške kažnjava instance u slučaju klasifikacije: potporni vektori i instance koje su dalje od njih u odgovarajućem poluprostoru ne doprinose grešci.
- Instance koje zađu u pojas doprinose grešci u skladu proporcionalno udaljenosti od hiperravni na kojoj leže potporni vektori.



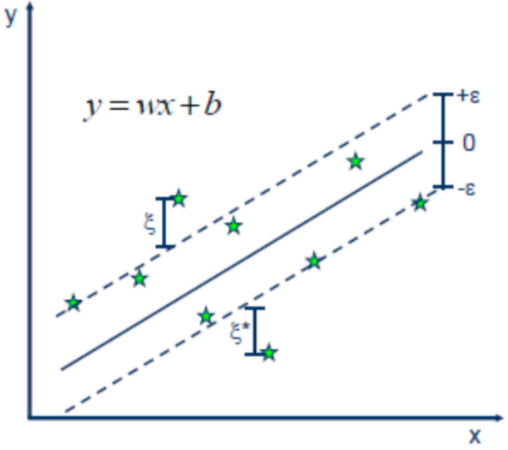
Široki pojas kod regresije

- U regresionom slučaju, tačke čija se vrednost razlikuje od vrednosti modela za manje od ε, ne doprinose grešci.



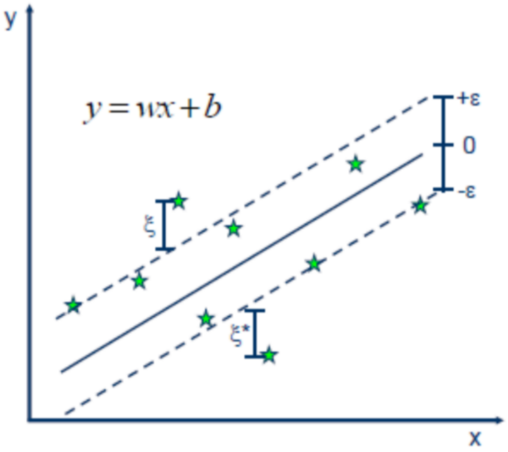
Široki pojas kod regresije

- U regresionom slučaju, tačke čija se vrednost razlikuje od vrednosti modela za manje od ε, ne doprinose grešci.
- Čim se razlikuju za više od ε, doprinose proporcionalno toj dodatnoj razlici.



Široki pojas kod regresije

- U regresionom slučaju, tačke čija se vrednost razlikuje od vrednosti modela za manje od ε, ne doprinose grešci.
- Čim se razlikuju za više od ε, doprinose proporcionalno toj dodatnoj razlici.
- To nas navodi na ideju da je u regresionom slučaju široki pojas zapravo prostor tačaka koje se po y osi razlikuju od regresione krive za više od ε.



Pregled

Metod potpornih vektora za klasifikaciju

Metod potpornih vektora za regresiju

Algoritam *k* najbližih suseda zasnovan na širokom pojasu

Algoritam k najbližih suseda zasnovan na širokom pojasu

- Jedan od najjednostavnijih algoritama mašinskog učenja

Algoritam *k* najbližih suseda zasnovan na širokom pojasu

- Jedan od najjednostavnijih algoritama mašinskog učenja
- Može služiti za klasifikaciju sa proizvoljnim brojem klasa, kao i za regresiju.

Algoritam k najbližih suseda zasnovan na širokom pojasu

- Jedan od najjednostavnijih algoritama mašinskog učenja
- Može služiti za klasifikaciju sa proizvoljnim brojem klasa, kao i za regresiju.
- Osnovna pretpostavka ovog algoritma je da nad podacima (nad njihovim vektorima atributa) možemo ustanoviti neku sličnost

Algoritam *k* najbližih suseda zasnovan na širokom pojasu

- Jedan od najjednostavnijih algoritama mašinskog učenja
- Može služiti za klasifikaciju sa proizvoljnim brojem klasa, kao i za regresiju.
- Osnovna pretpostavka ovog algoritma je da nad podacima (nad njihovim vektorima atributa) možemo ustanoviti neku sličnost
- Najčešće se pretpostavlja vektorska reprezentacija instanci i euklidsko rastojanje, ali moguće su i opštije pretpostavke.

Algoritam k najbližih suseda zasnovan na širokom pojasu

- Jedan od najjednostavnijih algoritama mašinskog učenja
- Može služiti za klasifikaciju sa proizvoljnim brojem klasa, kao i za regresiju.
- Osnovna pretpostavka ovog algoritma je da nad podacima (nad njihovim vektorima atributa) možemo ustanoviti neku sličnost
- Najčešće se pretpostavlja vektorska reprezentacija instanci i euklidsko rastojanje, ali moguće su i opštije pretpostavke.
- Ovaj algoritam retko predstavlja najbolji izbor za rešavanje nekog problema, ali neretko daje relativno dobre rezultate, a izuzetno lako se implementira i primenjuje

Algoritam k najbližih suseda zasnovan na širokom pojasu

- Kod klasifikacije: nepoznata instanca se klasifikuje tako što pronalazi k instanci iz skupa za obučavanje koje su joj najbliže u smislu neke izabrane metrike i pridružuje joj klasu koja se najčešće javlja među tih k instanci.

Algoritam k najbližih suseda zasnovan na širokom pojasu

- Kod klasifikacije: nepoznata instanca se klasifikuje tako što pronalazi k instanci iz skupa za obučavanje koje su joj najbliže u smislu neke izabrane metrike i pridružuje joj klasu koja se najčešće javlja među tih k instanci.
- Kod regresije: za predviđanje se uzima prosečna vrednost k najbližih suseda iz skupa za obučavanje.

Algoritam *k* najbližih suseda zasnovan na širokom pojasu

- Jedan od problema vezanih za algoritam *k* najbližih suseda je činjenica da se funkcija rastojanja bira *apriori*, nezavisno od podataka

Algoritam k najbližih suseda zasnovan na širokom pojasu

- Jedan od problema vezanih za algoritam k najbližih suseda je činjenica da se funkcija rastojanja bira *apriori*, nezavisno od podataka
- Na primer, ako se odabere euklidsko rastojanje, jednako će biti uzeti u obzir svi atributi a možda nisu svi jednako važni

Algoritam *k* najbližih suseda zasnovan na širokom pojasu

- Jedan od problema vezanih za algoritam *k* najbližih suseda je činjenica da se funkcija rastojanja bira *apriori*, nezavisno od podataka
- Na primer, ako se odabere euklidsko rastojanje, jednako će biti uzeti u obzir svi atributi a možda nisu svi jednako važni
- Dakle, potrebno je da funkciju rastojanja nekako zamenimo

Algoritam k najbližih suseda zasnovan na širokom pojasu

- Jedan od problema vezanih za algoritam k najbližih suseda je činjenica da se funkcija rastojanja bira *apriori*, nezavisno od podataka
- Na primer, ako se odabere euklidsko rastojanje, jednako će biti uzeti u obzir svi atributi a možda nisu svi jednako važni
- Dakle, potrebno je da funkciju rastojanja nekako zamenimo
- Jedno rešenje: u slučaju da domenski ekspert zna nešto više o svojstvima podataka, moguće je napraviti meru rastojanja koja će biti prilagođena datom problemu (pitanje je koliki domet bi imalo ovakvo rešenje)

Algoritam *k* najbližih suseda zasnovan na širokom pojasu

- Jedan od problema vezanih za algoritam k najbližih suseda je činjenica da se funkcija rastojanja bira *apriori*, nezavisno od podataka
- Na primer, ako se odabere euklidsko rastojanje, jednako će biti uzeti u obzir svi atributi a možda nisu svi jednako važni
- Dakle, potrebno je da funkciju rastojanja nekako zamenimo
- Jedno rešenje: u slučaju da domenski ekspert zna nešto više o svojstvima podataka, moguće je napraviti meru rastojanja koja će biti prilagođena datom problemu (pitanje je koliki domet bi imalo ovakvo rešenje)
- Bolje rešenje: naučiti funkciju rastojanja iz podataka

Funkcija rastojanja

- Ako neku funkciju treba da učimo, obično je parametrizujemo i naučimo parametre

Funkcija rastojanja

- Ako neku funkciju treba da učimo, obično je parametrizujemo i naučimo parametre
- Ukoliko je M pozitivno semidefinitna matrica (važi $x^T M x \geq 0$ za svako x), funkcija

$$d_M(x, x') = (x - x')^T M(x - x')$$

je funkcija rastojanja

Funkcija rastojanja

- Ako neku funkciju treba da učimo, obično je parametrizujemo i naučimo parametre
- Ukoliko je M pozitivno semidefinitna matrica (važi $x^T Mx \geq 0$ za svako x), funkcija
$$d_M(x, x') = (x - x')^T M(x - x')$$
je funkcija rastojanja
- Ako važi $M = I$, mesto tačaka jednakog rastojanja od neke fiksirane tačke C je sfera sa centrom u tački C (svodi se na obično euklidsko rastojanje)

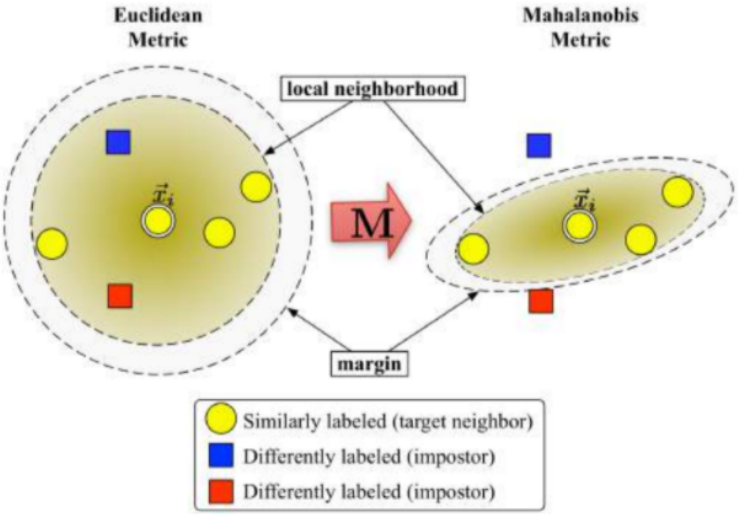
Funkcija rastojanja

- Ako neku funkciju treba da učimo, obično je parametrizujemo i naučimo parametre
- Ukoliko je M pozitivno semidefinitna matrica (važi $x^T M x \geq 0$ za svako x), funkcija
$$d_M(x, x') = (x - x')^T M (x - x')$$
je funkcija rastojanja
- Ako važi $M = I$, mesto tačaka jednakog rastojanja od neke fiksirane tačke C je sfera sa centrom u tački C (svodi se na obično euklidsko rastojanje)
- Ukoliko je matrica M dijagonalna (ne nužno jedinična), mesto tačaka jednakog rastojanja od C je elipsoid sa centrom u tački C, čije su ose paralelne koordinatnim osama

Funkcija rastojanja

- Ako neku funkciju treba da učimo, obično je parametrizujemo i naučimo parametre
- Ukoliko je M pozitivno semidefinitna matrica (važi $x^T Mx \geq 0$ za svako x), funkcija
$$d_M(x, x') = (x - x')^T M(x - x')$$
je funkcija rastojanja
- Ako važi $M = I$, mesto tačaka jednakog rastojanja od neke fiksirane tačke C je sfera sa centrom u tački C (svodi se na obično euklidsko rastojanje)
- Ukoliko je matrica M dijagonalna (ne nužno jedinična), mesto tačaka jednakog rastojanja od C je elipsoid sa centrom u tački C, čije su ose paralelne koordinatnim osama
- U opštem slučaju, radi se o proizvoljnom elipsoidu sa centrom u tački C (može biti proizvoljno zarotiran)

Funkcija rastojanja



Funkcija rastojanja

- Razmotrimo još jedan način razumevanja matrice M

Funkcija rastojanja

- Razmotrimo još jedan način razumevanja matrice M
- U opštem slučaju, kada imamo podatke, u procesu optimizacije ne možemo garantovati da će matrica M biti pozitivno semidefinitna

Funkcija rastojanja

- Razmotrimo još jedan način razumevanja matrice M
- U opštem slučaju, kada imamo podatke, u procesu optimizacije ne možemo garantovati da će matrica M biti pozitivno semidefinitna
- Postoji način kako da osiguramo da će uvek biti pozitivno semidefinitna

Funkcija rastojanja

- Razmotrimo još jedan način razumevanja matrice M
- U opštem slučaju, kada imamo podatke, u procesu optimizacije ne možemo garantovati da će matrica M biti pozitivno semidefinitna
- Postoji način kako da osiguramo da će uvek biti pozitivno semidefinitna
- Kako je matrica M pozitivno semidefinitna, može se predstaviti kao $M = Q^T Q$

Funkcija rastojanja

- Razmotrimo još jedan način razumevanja matrice M
- U opštem slučaju, kada imamo podatke, u procesu optimizacije ne možemo garantovati da će matrica M biti pozitivno semidefinitna
- Postoji način kako da osiguramo da će uvek biti pozitivno semidefinitna
- Kako je matrica M pozitivno semidefinitna, može se predstaviti kao $M = Q^T Q$
- Tada se metrika predstavlja kao

$$d_M(x, x') = (x - x')^T Q^T Q(x - x') = (Qx - Qx')^T (Qx - Qx')$$

Funkcija rastojanja

- Razmotrimo još jedan način razumevanja matrice M
- U opštem slučaju, kada imamo podatke, u procesu optimizacije ne možemo garantovati da će matrica M biti pozitivno semidefinitna
- Postoji način kako da osiguramo da će uvek biti pozitivno semidefinitna
- Kako je matrica M pozitivno semidefinitna, može se predstaviti kao $M = Q^T Q$
- Tada se metrika predstavlja kao

$$d_M(x, x') = (x - x')^T Q^T Q(x - x') = (Qx - Qx')^T (Qx - Qx')$$

- Zamenom koordinata $t = Qx$ dobija se

$$d_M(x, x') = (t - t')^T (t - t') = d_I(t, t')$$

Funkcija rastojanja

- Razmotrimo još jedan način razumevanja matrice M
- U opštem slučaju, kada imamo podatke, u procesu optimizacije ne možemo garantovati da će matrica M biti pozitivno semidefinitna
- Postoji način kako da osiguramo da će uvek biti pozitivno semidefinitna
- Kako je matrica M pozitivno semidefinitna, može se predstaviti kao $M = Q^T Q$
- Tada se metrika predstavlja kao

$$d_M(x, x') = (x - x')^T Q^T Q(x - x') = (Qx - Qx')^T (Qx - Qx')$$

- Zamenom koordinata $t = Qx$ dobija se

$$d_M(x, x') = (t - t')^T (t - t') = d_I(t, t')$$

- Drugim rečima, matrica Q transformiše prostor atributa tako da se metrika d_M u novim koordinatama može računati kao standardna euklidska metrika.

Funkcija rastojanja

- Umesto da se matrica M zada unapred, poželjno je učiti je iz podataka.

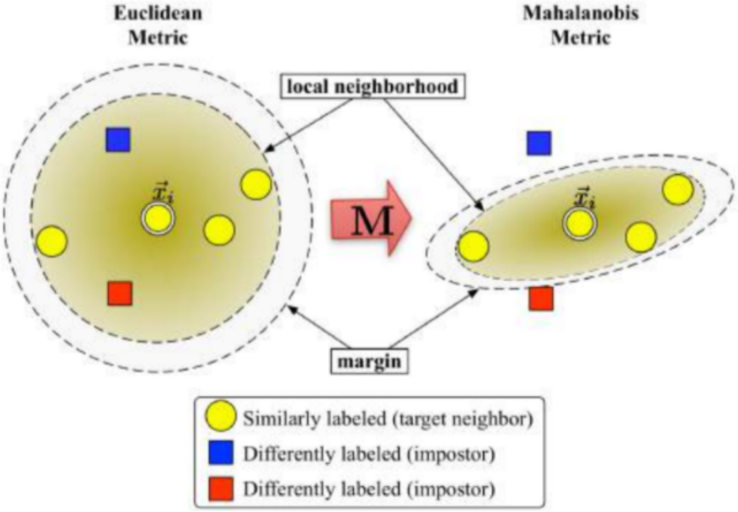
Funkcija rastojanja

- Umesto da se matrica M zada unapred, poželjno je učiti je iz podataka.
- Postavlja se pitanje kriterijuma u odnosu na koji se uči

Funkcija rastojanja

- Umesto da se matrica M zada unapred, poželjno je učiti je iz podataka.
- Postavlja se pitanje kriterijuma u odnosu na koji se uči
- Dobra matrica rastojanja bi bila ona za koju su tačke iz iste klase blizu, dok su sve tačke iz različitih klasa međusobno daleko.

Funkcija rastojanja



Određivanje matrice M

- Dobra matrica rastojanja bi bila ona za koju su tačke iz iste klase blizu, dok su sve tačke iz različitih klasa međusobno daleko.

Određivanje matrice M

- Dobra matrica rastojanja bi bila ona za koju su tačke iz iste klase blizu, dok su sve tačke iz različitih klasa međusobno daleko.
- Formalizacija: neka je

$$\mathcal{Z} = \{(x, x', x'') | (x, y), (x', y), (x'', y'') \in \mathcal{D} \land y \neq y''\}$$

skup svih trojki vektora atributa takvih da prva dva pripadaju istoj klasi, kojoj treći ne pripada

Određivanje matrice M

- Dobra matrica rastojanja bi bila ona za koju su tačke iz iste klase blizu, dok su sve tačke iz različitih klasa međusobno daleko.
- Formalizacija: neka je

$$\mathcal{Z} = \{(x, x', x'') | (x, y), (x', y), (x'', y'') \in \mathcal{D} \wedge y \neq y''\}$$

  skup svih trojki vektora atributa takvih da prva dva pripadaju istoj klasi, kojoj treći ne pripada
- Jedna formulacija metoda za određivanje matrice M bi mogla biti:

$$\min_{M} \sum_{(x,y),(x',y)\in\mathcal{D}} d_M(x, x')$$

$$d_M(x, x') < d_M(x, x'') \text{ za sve } (x, x', x'') \in \mathcal{Z}$$

$$M \succeq 0$$

  gde poslednji uslov označava pozitivnu semidefinitnost matrice M.

Određivanje matrice M

- Jedna formulacija metoda za određivanje matrice M bi mogla biti:

$$\min_{M} \sum_{(x,y),(x',y)\in\mathcal{D}} d_M(x,x')$$

$$d_M(x,x') < d_M(x,x'') \text{ za sve } (x,x',x'') \in \mathcal{Z}$$

$$M \succeq 0$$

Određivanje matrice M

- Jedna formulacija metoda za određivanje matrice M bi mogla biti:

$$\min_{M} \sum_{(x,y),(x',y)\in\mathcal{D}} d_M(x,x')$$

$$d_M(x,x') < d_M(x,x'') \text{ za sve } (x,x',x'') \in \mathcal{Z}$$

$$M \succeq 0$$

- Bolji pristup, koji insistira na postojanju širokog pojasa između elemenata klasa je sledeći:

$$\min_{M} \sum_{(x,y),(x',y)\in\mathcal{D}} d_M(x,x')$$

$$d_M(x,x') + 1 \leq d_M(x,x'') \text{ za sve } (x,x',x'') \in \mathcal{Z}$$

$$M \succeq 0$$

Određivanje matrice M

- Bolji pristup, koji insistira na postojanju širokog pojasa između elemenata klasa je sledeći:

$$\min_{M} \sum_{(x,y),(x',y)\in\mathcal{D}} d_M(x,x')$$

$$d_M(x,x') + 1 \leq d_M(x,x'') \text{ za sve } (x,x',x'') \in \mathcal{Z}$$

$$M \succeq 0$$

Određivanje matrice M

- Bolji pristup, koji insistira na postojanju širokog pojasa između elemenata klasa je sledeći:

$$\min_M \sum_{(x,y),(x',y)\in\mathcal{D}} d_M(x,x')$$
$$d_M(x,x') + 1 \leq d_M(x,x'') \text{ za sve } (x,x',x'') \in \mathcal{Z}$$
$$M \succeq 0$$

- Kao i u slučaju metoda potpornih vektora, krutost ograničenja se prevazilazi mekim pojasom, odnosno uvođenjem novih promenljivih koje čine model tolerantnijim na greške. Finalna formulacija glasi:

$$\min_M \sum_{(x,y),(x',y)\in\mathcal{D}} d_M(x,x') + C\sum_{i=1}^{|\mathcal{Z}|}\xi_i$$
$$d_M(x_i,x_i') + 1 \leq d_M(x_i,x_i'') + \xi_i \text{ za sve } (x_i,x_i',x_i'') \in \mathcal{Z}$$
$$\xi_i \geq 0 \qquad i = 1,\ldots,|\mathcal{Z}|$$
$$M \succeq 0$$