Regularizacija

Mašinsko učenje 2020/21.

Matematički fakultet
Univerzitet u Beogradu

Regularizacija

- Podešavanje prilagodljivosti modela

Regularizacija

- Podešavanje prilagodljivosti modela
- Nametanje specifične strukture modela

Regularizacija

- Podešavanje prilagodljivosti modela
- Nametanje specifične strukture modela
- Uključivanje domenskog znanja u model

Regularizacija

- ℓ_2 regularizacija

$$\Omega(w) = \|w\|_2^2 = \sum_{i=1}^{n} w_i^2$$

Regularizacija

- ℓ_2 regularizacija

$$\Omega(w) = \|w\|_2^2 = \sum_{i=1}^{n} w_i^2$$

- ℓ_1 regularizacija

$$\Omega(w) = \|w\|_1 = \sum_{i=1}^{n} |w_i|$$

Pregled

Proređeni modeli

Modeli složenije strukture i uključivanje domenskog znanja

Proređenost modela

- Model se smatra utoliko proređenijim ukoliko ima veći broj koeficijenata sa vrednošću nula

Proređenost modela

- Model se smatra utoliko proređenijim ukoliko ima veći broj koeficijenata sa vrednošću nula
- Proređenost modela znači da je njegovim obučavanjem obavljen posao *izbora atributa* (eng. *feature selection*)

Značaj proređenosti modela

Ukoliko je neki koeficijent modela nula, to znači:

- da postoji nešto u strukturi modela što nije bitno i što se ne mora izračunavati

Značaj proređenosti modela

Ukoliko je neki koeficijent modela nula, to znači:

- da postoji nešto u strukturi modela što nije bitno i što se ne mora izračunavati
- da neka od zavisnosti koje model može da izrazi ne postoji, odnosno da je model jednostavniji, što je generalno poželjno svojstvo modela u kontekstu moći generalizacije

Značaj proređenosti modela

Ukoliko je neki koeficijent modela nula, to znači:

- da postoji nešto u strukturi modela što nije bitno i što se ne mora izračunavati
- da neka od zavisnosti koje model može da izrazi ne postoji, odnosno da je model jednostavniji, što je generalno poželjno svojstvo modela u kontekstu moći generalizacije
- da model ima manje parametara koji je lakše analizirati i razumeti, odnosno proređeniji model je interpretabilniji

Laso regularizacija

- Proređeni modeli se često dobijaju tako što se neki metod učenja modifikuje posebnim vidom regularizacije – najčešće tako što se u minimizacionom problemu kao regularizacioni izraz upotrebi ℓ_1 norma:

$$\|w\|_1 = \sum_{i=1}^{n} |w_i|$$

Laso regularizacija

- Proređeni modeli se često dobijaju tako što se neki metod učenja modifikuje posebnim vidom regularizacije – najčešće tako što se u minimizacionom problemu kao regularizacioni izraz upotrebi ℓ_1 norma:

$$\|w\|_1 = \sum_{i=1}^{n} |w_i|$$

- Ovaj metod se naziva *laso* regularizacijom (eng. *lasso – least absolute shrinkage and selection operator*)

Laso regularizacija

Ključna razlika ℓ_1 regularizacije u odnosu na ℓ_2 regularizaciju:

- ℓ_2 regularizacija vodi smanjivanju apsolutnih vrednosti koeficijenata, često tako da veliki broj koeficijenata postane mali, ali i dalje različit od nule

Laso regularizacija

Ključna razlika ℓ_1 regularizacije u odnosu na ℓ_2 regularizaciju:

- ℓ_2 regularizacija vodi smanjivanju apsolutnih vrednosti koeficijenata, često tako da veliki broj koeficijenata postane mali, ali i dalje različit od nule
  - Kako izgleda ponašanje koeficijenata w kada koristimo ℓ_2?

Laso regularizacija

Ključna razlika ℓ_1 regularizacije u odnosu na ℓ_2 regularizaciju:

- ℓ_2 regularizacija vodi smanjivanju apsolutnih vrednosti koeficijenata, često tako da veliki broj koeficijenata postane mali, ali i dalje različit od nule
  - Kako izgleda ponašanje koeficijenata w kada koristimo ℓ_2?
  - Svi koeficijenti, neki brže, neki sporije, idu ka nuli, ali ne postaju tačno jednaki nuli do nekog poznog trenutka kada manje-više svi postaju jednaki nuli, a do tad se svi smanjuju

Laso regularizacija

Ključna razlika ℓ_1 regularizacije u odnosu na ℓ_2 regularizaciju:

- ℓ_2 regularizacija vodi smanjivanju apsolutnih vrednosti koeficijenata, često tako da veliki broj koeficijenata postane mali, ali i dalje različit od nule
  - Kako izgleda ponašanje koeficijenata w kada koristimo ℓ_2?
  - Svi koeficijenti, neki brže, neki sporije, idu ka nuli, ali ne postaju tačno jednaki nuli do nekog poznog trenutka kada manje-više svi postaju jednaki nuli, a do tad se svi smanjuju
- ℓ_1 regularizacija vodi tome da neki, manje važni koeficijenti postanu baš jednaki nuli

Laso regularizacija

Ključna razlika ℓ_1 regularizacije u odnosu na ℓ_2 regularizaciju:

- ℓ_2 regularizacija vodi smanjivanju apsolutnih vrednosti koeficijenata, često tako da veliki broj koeficijenata postane mali, ali i dalje različit od nule
  - Kako izgleda ponašanje koeficijenata w kada koristimo ℓ_2?
  - Svi koeficijenti, neki brže, neki sporije, idu ka nuli, ali ne postaju tačno jednaki nuli do nekog poznog trenutka kada manje-više svi postaju jednaki nuli, a do tad se svi smanjuju
- ℓ_1 regularizacija vodi tome da neki, manje važni koeficijenti postanu baš jednaki nuli
  - Kako izgleda ponašanje koeficijenata w kada koristimo ℓ_1?

Laso regularizacija

Ključna razlika ℓ_1 regularizacije u odnosu na ℓ_2 regularizaciju:

- ℓ_2 regularizacija vodi smanjivanju apsolutnih vrednosti koeficijenata, često tako da veliki broj koeficijenata postane mali, ali i dalje različit od nule
  - Kako izgleda ponašanje koeficijenata w kada koristimo ℓ_2?
  - Svi koeficijenti, neki brže, neki sporije, idu ka nuli, ali ne postaju tačno jednaki nuli do nekog poznog trenutka kada manje-više svi postaju jednaki nuli, a do tad se svi smanjuju
- ℓ_1 regularizacija vodi tome da neki, manje važni koeficijenti postanu baš jednaki nuli
  - Kako izgleda ponašanje koeficijenata w kada koristimo ℓ_1?
  - Neki koeficijenti postanu jednaki nuli relativno rano, u smislu kako povećavamo λ i rešavamo optimizacioni problem

Laso regularizacija

Ključna razlika ℓ_1 regularizacije u odnosu na ℓ_2 regularizaciju:

- ℓ_2 regularizacija vodi smanjivanju apsolutnih vrednosti koeficijenata, često tako da veliki broj koeficijenata postane mali, ali i dalje različit od nule
  - Kako izgleda ponašanje koeficijenata w kada koristimo ℓ_2?
  - Svi koeficijenti, neki brže, neki sporije, idu ka nuli, ali ne postaju tačno jednaki nuli do nekog poznog trenutka kada manje-više svi postaju jednaki nuli, a do tad se svi smanjuju
- ℓ_1 regularizacija vodi tome da neki, manje važni koeficijenti postanu baš jednaki nuli
  - Kako izgleda ponašanje koeficijenata w kada koristimo ℓ_1?
  - Neki koeficijenti postanu jednaki nuli relativno rano, u smislu kako povećavamo λ i rešavamo optimizacioni problem
  - Već za male vrednosti λ će neki koeficijenti postati tačno nula

Laso regularizacija

Ključna razlika ℓ_1 regularizacije u odnosu na ℓ_2 regularizaciju:

- ℓ_2 regularizacija vodi smanjivanju apsolutnih vrednosti koeficijenata, često tako da veliki broj koeficijenata postane mali, ali i dalje različit od nule
  - Kako izgleda ponašanje koeficijenata w kada koristimo ℓ_2?
  - Svi koeficijenti, neki brže, neki sporije, idu ka nuli, ali ne postaju tačno jednaki nuli do nekog poznog trenutka kada manje-više svi postaju jednaki nuli, a do tad se svi smanjuju
- ℓ_1 regularizacija vodi tome da neki, manje važni koeficijenti postanu baš jednaki nuli
  - Kako izgleda ponašanje koeficijenata w kada koristimo ℓ_1?
  - Neki koeficijenti postanu jednaki nuli relativno rano, u smislu kako povećavamo λ i rešavamo optimizacioni problem
  - Već za male vrednosti λ će neki koeficijenti postati tačno nula
- U specijalnom slučaju linearne regresije u kojem za matricu podataka važi $X^T X = I$, mogu se preciznije okarakterisati efekti ove dve regularizacije

Laso regularizacija

- Kako bismo preciznije okarakterisali efekte ove dve regularizacije, posmatrajmo specijalan slučaj kada su podaci ortogonalni

Laso regularizacija

- Kako bismo preciznije okarakterisali efekte ove dve regularizacije, posmatrajmo specijalan slučaj kada su podaci ortogonalni
- Kod ortogonalnih podataka, za matricu podataka važi $X^T X = I$

Laso regularizacija

- Kako bismo preciznije okarakterisali efekte ove dve regularizacije, posmatrajmo specijalan slučaj kada su podaci ortogonalni
- Kod ortogonalnih podataka, za matricu podataka važi $X^T X = I$
- Za podatke ovaj uslov retko važi ali se podaci mogu transformisati tako da ovaj uslov važi i onda možemo posmatrati takvu reprezentaciju

Laso regularizacija

- Posmatrajmo metodu linearne regresije. Neka je:

Laso regularizacija

- Posmatrajmo metodu linearne regresije. Neka je:
  - w vektor vrednosti parametara modela koji se dobija bez regularizacije

Laso regularizacija

- Posmatrajmo metodu linearne regresije. Neka je:
  - w vektor vrednosti parametara modela koji se dobija bez regularizacije
  - w' vektor vrednosti parametara koji se dobija pri ℓ_1 regularizaciji

Laso regularizacija

- Posmatrajmo metodu linearne regresije. Neka je:
  - w vektor vrednosti parametara modela koji se dobija bez regularizacije
  - w' vektor vrednosti parametara koji se dobija pri ℓ_1 regularizaciji
  - a w'' vektor vrednosti parametara koji se dobijaju pri ℓ_2 regularizaciji

Laso regularizacija

- Posmatrajmo metodu linearne regresije. Neka je:
  - w vektor vrednosti parametara modela koji se dobija bez regularizacije
  - w' vektor vrednosti parametara koji se dobija pri ℓ_1 regularizaciji
  - a w'' vektor vrednosti parametara koji se dobijaju pri ℓ_2 regularizaciji
  - λ vrednost regularizacionog parametra

Laso regularizacija

- Posmatrajmo metodu linearne regresije. Neka je:
  - w vektor vrednosti parametara modela koji se dobija bez regularizacije
  - w' vektor vrednosti parametara koji se dobija pri ℓ_1 regularizaciji
  - a w'' vektor vrednosti parametara koji se dobijaju pri ℓ_2 regularizaciji
  - λ vrednost regularizacionog parametra
  - skup podataka je ortogonalan odnosno važi $X^T X = I$

Laso regularizacija

- Posmatrajmo metodu linearne regresije. Neka je:
  - w vektor vrednosti parametara modela koji se dobija bez regularizacije
  - w' vektor vrednosti parametara koji se dobija pri ℓ_1 regularizaciji
  - a w'' vektor vrednosti parametara koji se dobijaju pri ℓ_2 regularizaciji
  - λ vrednost regularizacionog parametra
  - skup podataka je ortogonalan odnosno važi $X^T X = I$
- Pod datim pretpostavkama, ispostavlja se da su i w' i w'' u nekom odnosu sa w:

$$w'_i = \mathrm{sgn}(w_i) \max\left(|w_i| - \frac{\lambda}{2}, 0\right) \qquad w''_i = \frac{w_i}{1+\lambda} \qquad i = 1, \ldots, n$$

Laso regularizacija

$$w_i' = \operatorname{sgn}(w_i)\max\left(|w_i| - \frac{\lambda}{2}, 0\right) \qquad w_i'' = \frac{w_i}{1+\lambda} \qquad i = 1, \ldots, n$$

Posmatrajmo izraz w_i''

- kako povećavamo λ, tako se w_i deli sve većim i većim brojem i dobijamo sve manji i manji broj

Laso regularizacija

$$w_i' = \mathrm{sgn}(w_i)\max\left(|w_i| - \frac{\lambda}{2}, 0\right) \qquad w_i'' = \frac{w_i}{1+\lambda} \qquad i = 1, \dots, n$$

Posmatrajmo izraz w_i''

- kako povećavamo λ, tako se w_i deli sve većim i većim brojem i dobijamo sve manji i manji broj
- w_i'' nikada neće postati nula (osim zbog potkoračenja)

Laso regularizacija

$$w_i' = \mathrm{sgn}(w_i) \max\left(|w_i| - \frac{\lambda}{2}, 0\right) \qquad w_i'' = \frac{w_i}{1+\lambda} \qquad i = 1, \ldots, n$$

Posmatrajmo izraz w_i'

- Ukoliko je λ malo, koeficijent ostaje istog znaka, ali apsolutne vrednosti umanjene za $\lambda/2$

Laso regularizacija

$$w_i' = \operatorname{sgn}(w_i)\max\left(|w_i| - \frac{\lambda}{2}, 0\right) \qquad w_i'' = \frac{w_i}{1+\lambda} \qquad i = 1, \ldots, n$$

Posmatrajmo izraz w_i'

- Ukoliko je λ malo, koeficijent ostaje istog znaka, ali apsolutne vrednosti umanjene za $\lambda/2$
- Ukoliko je λ dovoljno veliko da $|w_i| - \lambda/2$ postane negativno, koeficijent će imati vrednosti nula

Laso regularizacija

$$w_i' = \operatorname{sgn}(w_i) \max\left(|w_i| - \frac{\lambda}{2}, 0\right) \qquad w_i'' = \frac{w_i}{1+\lambda} \qquad i = 1, \ldots, n$$

- Kod ℓ_1 regularizacije, oduzimanjem konačne, dovoljno velike vrednosti λ možemo doći do koeficijenta vrednosti 0

Laso regularizacija

$$w_i' = \operatorname{sgn}(w_i) \max\left(|w_i| - \frac{\lambda}{2}, 0\right) \qquad w_i'' = \frac{w_i}{1+\lambda} \qquad i = 1, \ldots, n$$

- Kod ℓ_1 regularizacije, oduzimanjem konačne, dovoljno velike vrednosti λ možemo doći do koeficijenta vrednosti 0
- Kod ℓ_2 regularizacije, deljenjem sa proizvoljno velikom, konačnom vrednošću λ ne možemo doći do koeficijenta vrednosti 0

Laso regularizacija

$$w_i' = \text{sgn}(w_i)\max\left(|w_i| - \frac{\lambda}{2}, 0\right) \qquad w_i'' = \frac{w_i}{1+\lambda} \qquad i = 1, \ldots, n$$

- Kod ℓ_1 regularizacije, oduzimanjem konačne, dovoljno velike vrednosti λ možemo doći do koeficijenta vrednosti 0
- Kod ℓ_2 regularizacije, deljenjem sa proizvoljno velikom, konačnom vrednošću λ ne možemo doći do koeficijenta vrednosti 0
- Nije očigledno da ovo zaista važi i da ℓ_1 zaista daje proređene modele

Laso regularizacija

$$w_i' = \operatorname{sgn}(w_i) \max\left(|w_i| - \frac{\lambda}{2}, 0\right) \qquad w_i'' = \frac{w_i}{1+\lambda} \qquad i = 1, \ldots, n$$

- Kod ℓ_1 regularizacije, oduzimanjem konačne, dovoljno velike vrednosti λ možemo doći do koeficijenta vrednosti 0
- Kod ℓ_2 regularizacije, deljenjem sa proizvoljno velikom, konačnom vrednošću λ ne možemo doći do koeficijenta vrednosti 0
- Nije očigledno da ovo zaista važi i da ℓ_1 zaista daje proređene modele
- Stoga ćemo dati još ilustracija

Laso regularizacija

- Svaki problem oblika

$$\min_w E(w, \mathcal{D}) + \lambda \|w\|_q$$

za $q \in \mathbb{N}^+$ ekvivalentan je problemu oblika

$$\min_w E(w, \mathcal{D})$$

$$\|w\|_q \leq t$$

Laso regularizacija

- Svaki problem oblika

$$\min_w E(w, \mathcal{D}) + \lambda \|w\|_q$$

za $q \in \mathbb{N}^+$ ekvivalentan je problemu oblika

$$\min_w E(w, \mathcal{D})$$

$$\|w\|_q \leq t$$

- To znači da za svaku vrednost metaparametra λ postoji neka vrednost metaparametra t, tako da je rešenje oba problema isto i obratno

Laso regularizacija

- Svaki problem oblika

$$\min_{w} E(w, \mathcal{D}) + \lambda \|w\|_q$$

za $q \in \mathbb{N}^+$ ekvivalentan je problemu oblika

$$\min_{w} E(w, \mathcal{D})$$

$$\|w\|_q \leq t$$

- To znači da za svaku vrednost metaparametra λ postoji neka vrednost metaparametra t, tako da je rešenje oba problema isto i obratno
- Postoji 1-na-1 korespodencija između λ i t (nisu jednaki)

Laso regularizacija

- Svaki problem oblika

$$\min_{w} E(w, \mathcal{D}) + \lambda \|w\|_q$$

za $q \in \mathbb{N}^+$ ekvivalentan je problemu oblika

$$\min_{w} E(w, \mathcal{D})$$

$$\|w\|_q \leq t$$

- To znači da za svaku vrednost metaparametra λ postoji neka vrednost metaparametra t, tako da je rešenje oba problema isto i obratno
- Postoji 1-na-1 korespodencija između λ i t (nisu jednaki)
- λ i t su u obrnutoj vezi - veća λ odgovara manjem t i obrnuto

Laso regularizacija

- U daljem razmatranju ćemo umesto optimizacionog problema bez ograničenja posmatrati optimizacioni problem sa ograničenjima:

$$\min_{w} E(w, \mathcal{D})$$

$$\|w\|_q \leq t$$

Laso regularizacija

- U daljem razmatranju ćemo umesto optimizacionog problema bez ograničenja posmatrati optimizacioni problem sa ograničenjima:

$$\min_{w} E(w, \mathcal{D})$$

$$\|w\|_q \leq t$$

- Hoćemo da minimizujemo funkciju greške pod uslovom da je naše rešenje iz neke lopte poluprečnika t

Laso regularizacija

- U daljem razmatranju ćemo umesto optimizacionog problema bez ograničenja posmatrati optimizacioni problem sa ograničenjima:

$$\min_{w} E(w, \mathcal{D})$$

$$\|w\|_q \leq t$$

- Hoćemo da minimizujemo funkciju greške pod uslovom da je naše rešenje iz neke lopte poluprečnika t
- Ako je to ℓ_2 norma, onda je to euklidska lopta

Laso regularizacija

- U daljem razmatranju ćemo umesto optimizacionog problema bez ograničenja posmatrati optimizacioni problem sa ograničenjima:

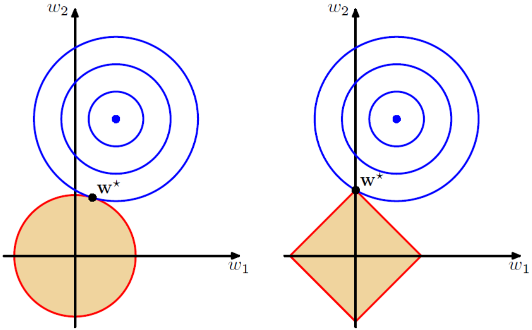
$$\min_{w} E(w, \mathcal{D})$$

$$\|w\|_q \leq t$$

- Hoćemo da minimizujemo funkciju greške pod uslovom da je naše rešenje iz neke lopte poluprečnika t
- Ako je to ℓ_2 norma, onda je to euklidska lopta
- Ako je to ℓ_1 norma, onda je to malo drugačija lopta

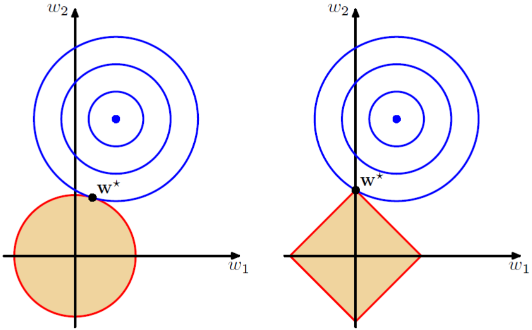
Laso regularizacija

- Posmatrajmo funkciju greške E i ograničenja $\|w\|_1 \leq t$ i $\|w\|_2 \leq t$ u dve dimenzije



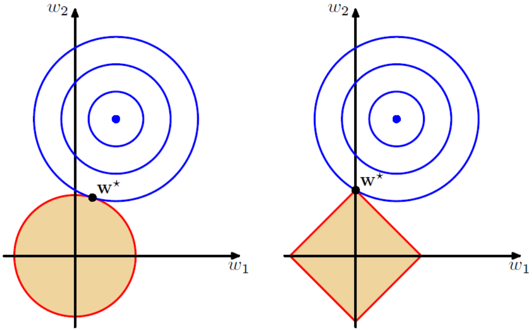
Laso regularizacija

- Posmatrajmo funkciju greške E i ograničenja $\|w\|_1 \leq t$ i $\|w\|_2 \leq t$ u dve dimenzije
- Funkcija greške je predstavljena konturama



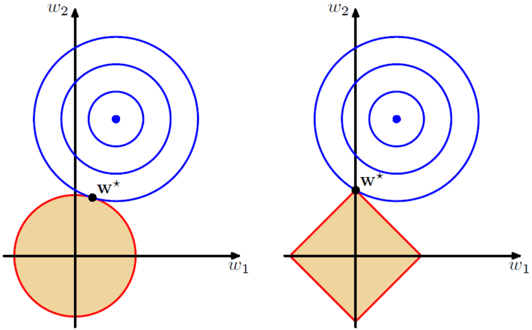
Laso regularizacija

- Posmatrajmo funkciju greške E i ograničenja $\|w\|_1 \leq t$ i $\|w\|_2 \leq t$ u dve dimenzije
- Funkcija greške je predstavljena konturama
  - u centru je minimum



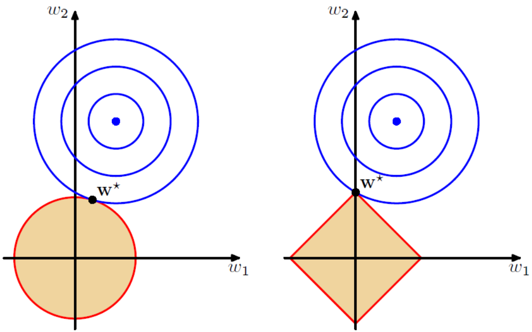
Laso regularizacija

- Posmatrajmo funkciju greške E i ograničenja $\|w\|_1 \leq t$ i $\|w\|_2 \leq t$ u dve dimenzije
- Funkcija greške je predstavljena konturama
  - u centru je minimum
  - tačke na istoj kružnici imaju istu vrednost greške



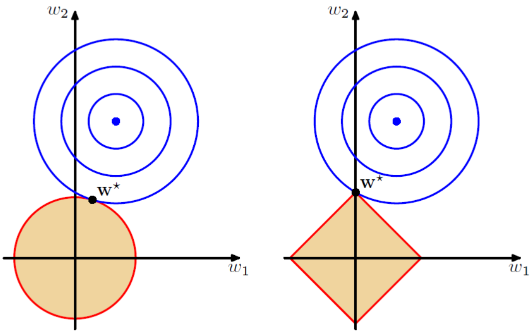
Laso regularizacija

- Posmatrajmo funkciju greške E i ograničenja $\|w\|_1 \leq t$ i $\|w\|_2 \leq t$ u dve dimenzije
- Funkcija greške je predstavljena konturama
  - u centru je minimum
  - tačke na istoj kružnici imaju istu vrednost greške
  - kako se udaljavamo od minimuma, povećava se vrednost greške



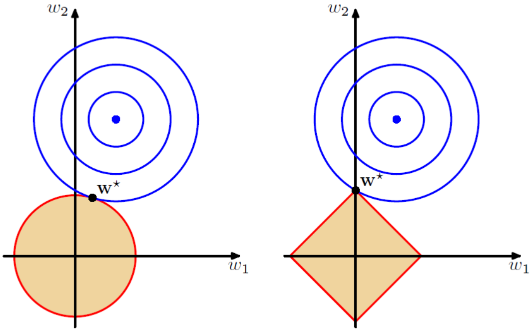
Laso regularizacija

- Posmatrajmo funkciju greške E i ograničenja $\|w\|_1 \leq t$ i $\|w\|_2 \leq t$ u dve dimenzije
- Funkcija greške je predstavljena konturama
  - u centru je minimum
  - tačke na istoj kružnici imaju istu vrednost greške
  - kako se udaljavamo od minimuma, povećava se vrednost greške
- Potrebno je da nađemo rešenje koje minimizuje funkciju greške i ispunjava ograničenja (pripada datoj lopti)



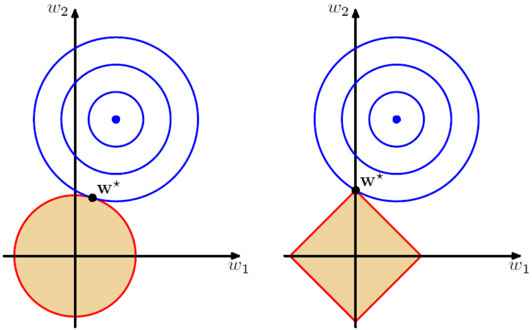
Laso regularizacija

- Potrebno je da nađemo rešenje koje minimizuje funkciju greške i ispunjava ograničenja (pripada datoj lopti)



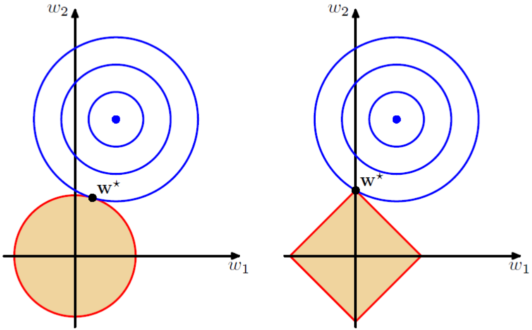
Laso regularizacija

- Potrebno je da nađemo rešenje koje minimizuje funkciju greške i ispunjava ograničenja (pripada datoj lopti)
- Kada ne bismo imali regularizaciju (kada ne bi bilo ograničenja da rešenje pripada lopti), nekim metodom optimizacije bismo pronašli minimalnu vrednost funkcije (centar)



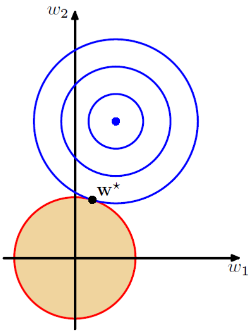
Laso regularizacija

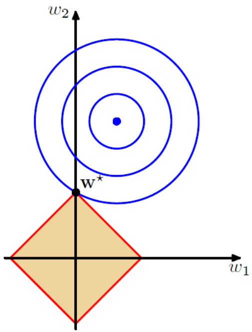
- Potrebno je da nađemo rešenje koje minimizuje funkciju greške i ispunjava ograničenja (pripada datoj lopti)
- Kada ne bismo imali regularizaciju (kada ne bi bilo ograničenja da rešenje pripada lopti), nekim metodom optimizacije bismo pronašli minimalnu vrednost funkcije (centar)
- Pošto imamo regularizaciju i rešenje mora da pripada lopti, rešenje je ona tačka na lopti koja je najbliža minimumu



Laso regularizacija

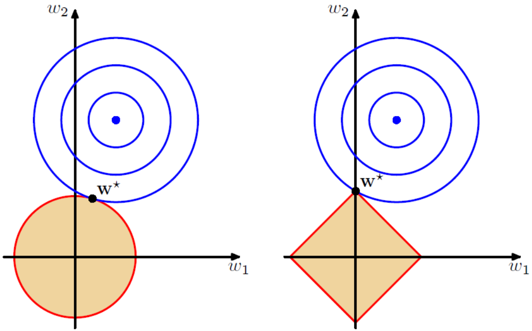
- Potrebno je da nađemo rešenje koje minimizuje funkciju greške i ispunjava ograničenja (pripada datoj lopti)





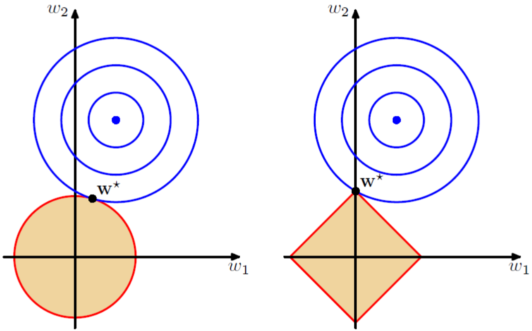
Laso regularizacija

- kod optimalnog rešenja u slučaju ℓ_2 regularizacije, oba koeficijenta su različita od nule



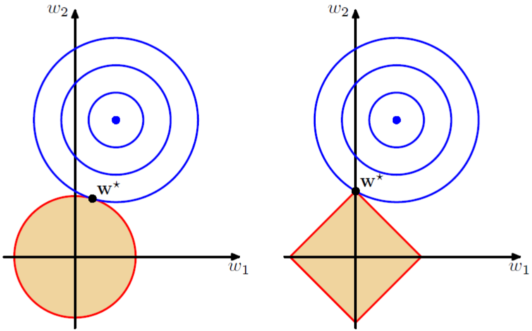
Laso regularizacija

- kod optimalnog rešenja u slučaju ℓ_2 regularizacije, oba koeficijenta su različita od nule
- kod optimalnog rešenja u slučaju ℓ_1 regularizacije, jedan koeficijent je jednak nuli a drugi je različit od nule



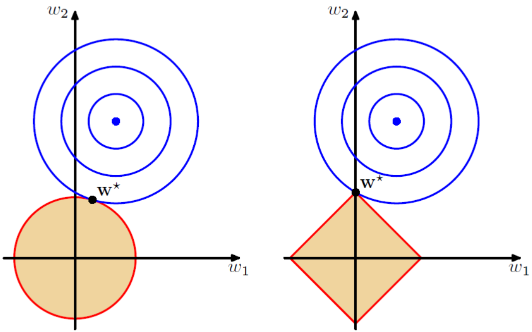
Laso regularizacija

- kod optimalnog rešenja u slučaju ℓ_2 regularizacije, oba koeficijenta su različita od nule
- kod optimalnog rešenja u slučaju ℓ_1 regularizacije, jedan koeficijent je jednak nuli a drugi je različit od nule
- zahvaljujući špicastom obliku ℓ_1 lopte, optimalno rešenje u slučaju ℓ_1 regularizacije ima 50% nenula parametara, dok optimalno rešenje u slučaju ℓ_2 lopte ima 100% nenula parametara



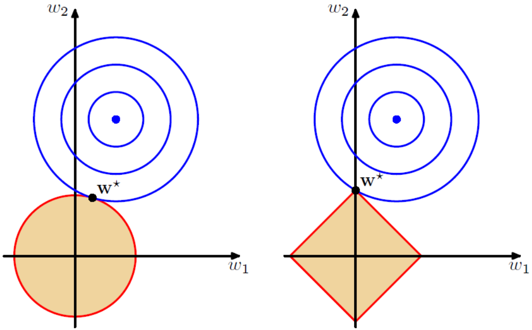
Laso regularizacija

- kod ℓ_1 regularizacije, nekada ne možemo dobiti proređeni model ma koliko povećavali λ



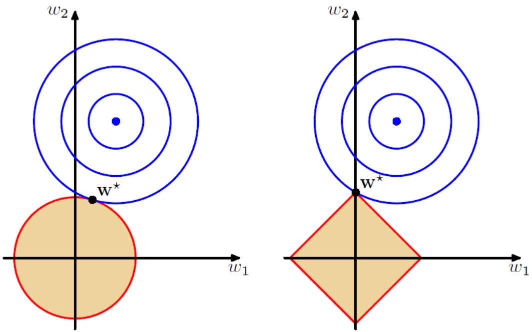
Laso regularizacija

- kod ℓ_1 regularizacije, nekada ne možemo dobiti proređeni model ma koliko povećavali λ
- sa povećanjem λ se smanjuje t a time i lopta u kojoj rešenje treba da se nalazi



Laso regularizacija

- kod ℓ_1 regularizacije, nekada ne možemo dobiti proređeni model ma koliko povećavali λ
- sa povećanjem λ se smanjuje t a time i lopta u kojoj rešenje treba da se nalazi
- ako je funkcija greške pomerena, ma koliko da je lopta mala, rešenje neće moći da bude na jednoj od osa



Laso regularizacija

- Još jedan pogled koji doprinosi razumevanju zašto ℓ_1 regularizacija vodi proređenim modelima, a ℓ_2 ne, je vezan za gradijente ovih normi

Laso regularizacija

- Još jedan pogled koji doprinosi razumevanju zašto ℓ_1 regularizacija vodi proređenim modelima, a ℓ_2 ne, je vezan za gradijente ovih normi
- Optimizacioni metodi obično počivaju na uzastopnom umanjivanju tekućih parametara modela za neki vektor proporcionalan gradijentu ciljne funkcije

Laso regularizacija

- Još jedan pogled koji doprinosi razumevanju zašto ℓ_1 regularizacija vodi proređenim modelima, a ℓ_2 ne, je vezan za gradijente ovih normi
- Optimizacioni metodi obično počivaju na uzastopnom umanjivanju tekućih parametara modela za neki vektor proporcionalan gradijentu ciljne funkcije
- U slučaju ℓ_1 regularizacije, doprinos regularizacionog izraza gradijentu ciljne funkcije je

$$\frac{\partial \|w\|_1}{\partial w_i} = \text{sgn}(w_i)$$

kad god važi $w_i \neq 0$. U slučaju ℓ_2 regularizacije važi

$$\frac{\partial \|w\|_2^2}{\partial w_i} = 2w_i$$

Laso regularizacija

- Još jedan pogled koji doprinosi razumevanju zašto ℓ_1 regularizacija vodi proređenim modelima, a ℓ_2 ne, je vezan za gradijente ovih normi
- Optimizacioni metodi obično počivaju na uzastopnom umanjivanju tekućih parametara modela za neki vektor proporcionalan gradijentu ciljne funkcije
- U slučaju ℓ_1 regularizacije, doprinos regularizacionog izraza gradijentu ciljne funkcije je

$$\frac{\partial \|w\|_1}{\partial w_i} = \mathrm{sgn}(w_i)$$

kad god važi $w_i \neq 0$. U slučaju ℓ_2 regularizacije važi

$$\frac{\partial \|w\|_2^2}{\partial w_i} = 2w_i$$

- Očito, kolika god da je vrednost parametra w_i, ℓ_1 regularizacija podjednako doprinosi smanjenju apsolutne vrednosti koeficijenta w_i. S druge strane, ℓ_2 regularizacija doprinosi utoliko manje, što je vrednost parametra manja, tako da kako se vrednost bliži nuli, regularizacija sve manje pomaže u daljem smanjenju

Laso regularizacija

- Očigledan problem vezan za ℓ_1 regularizaciju je njena nediferencijabilnost

Laso regularizacija

- Očigledan problem vezan za ℓ_1 regularizaciju je njena nediferencijabilnost
- Nediferencijabilnost se u mašinskom učenju često ignoriše, pošto u nekim problemima nije mnogo verovatno da će optimizacioni algoritam zasnovan na gradijentima naleteti na tačku u kojoj je funkcija nediferencijabilna

Laso regularizacija

- Očigledan problem vezan za ℓ_1 regularizaciju je njena nediferencijabilnost
- Nediferencijabilnost se u mašinskom učenju često ignoriše, pošto u nekim problemima nije mnogo verovatno da će optimizacioni algoritam zasnovan na gradijentima naleteti na tačku u kojoj je funkcija nediferencijabilna
- Čak i ako naleti i preduzme korak u pogrešnom pravcu, taj korak će u mnogim slučajevima biti kompenzovan daljim tokom optimizacije

Laso regularizacija

- Očigledan problem vezan za ℓ_1 regularizaciju je njena nediferencijabilnost
- Nediferencijabilnost se u mašinskom učenju često ignoriše, pošto u nekim problemima nije mnogo verovatno da će optimizacioni algoritam zasnovan na gradijentima naleteti na tačku u kojoj je funkcija nediferencijabilna
- Čak i ako naleti i preduzme korak u pogrešnom pravcu, taj korak će u mnogim slučajevima biti kompenzovan daljim tokom optimizacije
- Ipak, u slučaju ℓ_1 regularizacije, ovaj problem se ne može ignorisati

Laso regularizacija

- Očigledan problem vezan za ℓ_1 regularizaciju je njena nediferencijabilnost
- Nediferencijabilnost se u mašinskom učenju često ignoriše, pošto u nekim problemima nije mnogo verovatno da će optimizacioni algoritam zasnovan na gradijentima naleteti na tačku u kojoj je funkcija nediferencijabilna
- Čak i ako naleti i preduzme korak u pogrešnom pravcu, taj korak će u mnogim slučajevima biti kompenzovan daljim tokom optimizacije
- Ipak, u slučaju ℓ_1 regularizacije, ovaj problem se ne može ignorisati
- Naime, tačka u kojoj je funkcija nediferencijabilna je baš tačka rešenja kojoj se teži, a ne neka tačka na koju optimizacija može nabasati, a u koju se ne mora vraćati

Laso regularizacija

- Očigledan problem vezan za ℓ_1 regularizaciju je njena nediferencijabilnost
- Nediferencijabilnost se u mašinskom učenju često ignoriše, pošto u nekim problemima nije mnogo verovatno da će optimizacioni algoritam zasnovan na gradijentima naleteti na tačku u kojoj je funkcija nediferencijabilna
- Čak i ako naleti i preduzme korak u pogrešnom pravcu, taj korak će u mnogim slučajevima biti kompenzovan daljim tokom optimizacije
- Ipak, u slučaju ℓ_1 regularizacije, ovaj problem se ne može ignorisati
- Naime, tačka u kojoj je funkcija nediferencijabilna je baš tačka rešenja kojoj se teži, a ne neka tačka na koju optimizacija može nabasati, a u koju se ne mora vraćati
- Otud se uz ovu vrstu regularizacije koriste i posebni optimizacioni algoritmi

Laso regularizacija

- Još jedna mana laso regularizacije je ta što vodi nestabilnim rešenjima

Laso regularizacija

- Još jedna mana laso regularizacije je ta što vodi nestabilnim rešenjima
- Naime, ukoliko postoje dva jako korelirana atributa, laso regularizacija će verovatno iz modela isključiti jedan od tih atributa

Laso regularizacija

- Još jedna mana laso regularizacije je ta što vodi nestabilnim rešenjima
- Naime, ukoliko postoje dva jako korelirana atributa, laso regularizacija će verovatno iz modela isključiti jedan od tih atributa
- Kako su atributi jako korelirani, nije velika razlika da li će biti izbačen jedan ili drugi i lako se može desiti da male promene u podacima vode isključivanju različitih atributa

Laso regularizacija

- Još jedna mana laso regularizacije je ta što vodi nestabilnim rešenjima
- Naime, ukoliko postoje dva jako korelirana atributa, laso regularizacija će verovatno iz modela isključiti jedan od tih atributa
- Kako su atributi jako korelirani, nije velika razlika da li će biti izbačen jedan ili drugi i lako se može desiti da male promene u podacima vode isključivanju različitih atributa
- Ovo očito predstavlja problem za interpretaciju modela

Laso regularizacija

- Još jedna mana laso regularizacije je ta što vodi nestabilnim rešenjima
- Naime, ukoliko postoje dva jako korelirana atributa, laso regularizacija će verovatno iz modela isključiti jedan od tih atributa
- Kako su atributi jako korelirani, nije velika razlika da li će biti izbačen jedan ili drugi i lako se može desiti da male promene u podacima vode isključivanju različitih atributa
- Ovo očito predstavlja problem za interpretaciju modela
- Jedno rešenje ovog problema je korišćenje *elastične mreže* (eng. *elastic net*), što je regularizacija kojoj odgovara izraz

$$\Omega(w) = \mu\|w\|_1 + (1-\mu)\|w\|_2^2$$

pri čemu važi $\mu \in [0, 1]$.

Pregled

Prorеđeni modeli

Modeli složenije strukture i uključivanje domenskog znanja

Modeli složenije strukture i uključivanje domenskog znanja

- Želimo da predvidimo prinos pšenice u Srbiji na osnovu nekih atributa

Modeli složenije strukture i uključivanje domenskog znanja

- Želimo da predvidimo prinos pšenice u Srbiji na osnovu nekih atributa
- Problem je što se u Srbiji tek od skora sakupljaju podaci koji se tiču prinosa i koji bi se mogli upotrebiti kao atributi: količina padavina, biljne bolesti i slično

Modeli složenije strukture i uključivanje domenskog znanja

- Želimo da predvidimo prinos pšenice u Srbiji na osnovu nekih atributa
- Problem je što se u Srbiji tek od skora sakupljaju podaci koji se tiču prinosa i koji bi se mogli upotrebiti kao atributi: količina padavina, biljne bolesti i slično
- Dakle, na raspolaganju nam je mala količina podataka

Modeli složenije strukture i uključivanje domenskog znanja

- Želimo da predvidimo prinos pšenice u Srbiji na osnovu nekih atributa
- Problem je što se u Srbiji tek od skora sakupljaju podaci koji se tiču prinosa i koji bi se mogli upotrebiti kao atributi: količina padavina, biljne bolesti i slično
- Dakle, na raspolaganju nam je mala količina podataka
- Neke druge zemlje ovakve podatke sakupljaju duže i imaju veće količine podataka i na osnovu njih razvijane javno dostupne modele koji predviđaju prinos pšenice

Modeli složenije strukture i uključivanje domenskog znanja

- Želimo da predvidimo prinos pšenice u Srbiji na osnovu nekih atributa
- Problem je što se u Srbiji tek od skora sakupljaju podaci koji se tiču prinosa i koji bi se mogli upotrebiti kao atributi: količina padavina, biljne bolesti i slično
- Dakle, na raspolaganju nam je mala količina podataka
- Neke druge zemlje ovakve podatke sakupljaju duže i imaju veće količine podataka i na osnovu njih razvijane javno dostupne modele koji predviđaju prinos pšenice
- Pretpostavimo da nam je poznat jedan takav model, na primer za Francusku

Modeli složenije strukture i uključivanje domenskog znanja

- Želimo da predvidimo prinos pšenice u Srbiji na osnovu nekih atributa
- Problem je što se u Srbiji tek od skora sakupljaju podaci koji se tiču prinosa i koji bi se mogli upotrebiti kao atributi: količina padavina, biljne bolesti i slično
- Dakle, na raspolaganju nam je mala količina podataka
- Neke druge zemlje ovakve podatke sakupljaju duže i imaju veće količine podataka i na osnovu njih razvijane javno dostupne modele koji predviđaju prinos pšenice
- Pretpostavimo da nam je poznat jedan takav model, na primer za Francusku
- Modeli za Srbiju i za Francusku verovatno neće biti isti, ali mogu biti vrlo slični jer se radi o sličnom podneblju i o istoj poljoprivrednoj kulturi

Modeli složenije strukture i uključivanje domenskog znanja

- Ako mi imamo malo podataka a Francuzi imaju puno, i dostupan nam je model za Francusku, da li možemo da nekako upotrebimo taj njihov model sa jedne strane i malu količinu sopstvenih podataka sa druge strane prilikom izrade modela za Srbiju?

Modeli složenije strukture i uključivanje domenskog znanja

- Ako mi imamo malo podataka a Francuzi imaju puno, i dostupan nam je model za Francusku, da li možemo da nekako upotrebimo taj njihov model sa jedne strane i malu količinu sopstvenih podataka sa druge strane prilikom izrade modela za Srbiju?
- Pritom, ne želimo da direktno primenimo francuski model već želimo i da uzmemo u obzir i sopstvene podatke

Modeli složenije strukture i uključivanje domenskog znanja

- Ako mi imamo malo podataka a Francuzi imaju puno, i dostupan nam je model za Francusku, da li možemo da nekako upotrebimo taj njihov model sa jedne strane i malu količinu sopstvenih podataka sa druge strane prilikom izrade modela za Srbiju?
- Pritom, ne želimo da direktno primenimo francuski model već želimo i da uzmemo u obzir i sopstvene podatke
- Želimo da napravimo balans između apriornog znanja koje nam nudi francuski model i znanja koje imamo u sopstvenim podacima

Modeli složenije strukture i uključivanje domenskog znanja

- Ako mi imamo malo podataka a Francuzi imaju puno, i dostupan nam je model za Francusku, da li možemo da nekako upotrebimo taj njihov model sa jedne strane i malu količinu sopstvenih podataka sa druge strane prilikom izrade modela za Srbiju?
- Pritom, ne želimo da direktno primenimo francuski model već želimo i da uzmemo u obzir i sopstvene podatke
- Želimo da napravimo balans između apriornog znanja koje nam nudi francuski model i znanja koje imamo u sopstvenim podacima
- Ovo predstavlja vid uključivanja prethodnog znanja u proces učenja.

Modeli složenije strukture i uključivanje domenskog znanja

- Pretpostavimo da modeliramo prinos pšenice u Srbiji regularizovanom linearnom regresijom:

$$\min_{w} \sum_{i=1}^{N} (y_i - w \cdot x_i)^2 + \lambda \Omega(w)$$

Modeli složenije strukture i uključivanje domenskog znanja

- Pretpostavimo da modeliramo prinos pšenice u Srbiji regularizovanom linearnom regresijom:

$$\min_{w} \sum_{i=1}^{N} (y_i - w \cdot x_i)^2 + \lambda \Omega(w)$$

- Uz to, pretpostavimo da imamo na raspolaganju francuski model (koeficijente w' tog modela) i francuski skup podataka

Modeli složenije strukture i uključivanje domenskog znanja

- Pretpostavimo da modeliramo prinos pšenice u Srbiji regularizovanom linearnom regresijom:

$$\min_{w} \sum_{i=1}^{N} (y_i - w \cdot x_i)^2 + \lambda \Omega(w)$$

- Uz to, pretpostavimo da imamo na raspolaganju francuski model (koeficijente w' tog modela) i francuski skup podataka
- Skupovi podataka za Francusku i Srbiju ne moraju imati iste atribute

Modeli složenije strukture i uključivanje domenskog znanja

- Pretpostavimo da modeliramo prinos pšenice u Srbiji regularizovanom linearnom regresijom:

$$\min_{w} \sum_{i=1}^{N} (y_i - w \cdot x_i)^2 + \lambda \Omega(w)$$

- Uz to, pretpostavimo da imamo na raspolaganju francuski model (koeficijente w' tog modela) i francuski skup podataka
- Skupovi podataka za Francusku i Srbiju ne moraju imati iste atribute
- Neka je G skup atributa koji je zajednički i za Francusku i za Srbiju

Modeli složenije strukture i uključivanje domenskog znanja

- Pretpostavimo da modeliramo prinos pšenice u Srbiji regularizovanom linearnom regresijom:

$$\min_w \sum_{i=1}^{N} (y_i - w \cdot x_i)^2 + \lambda\Omega(w)$$

- Uz to, pretpostavimo da imamo na raspolaganju francuski model (koeficijente w' tog modela) i francuski skup podataka
- Skupovi podataka za Francusku i Srbiju ne moraju imati iste atribute
- Neka je G skup atributa koji je zajednički i za Francusku i za Srbiju
- Neka je w_G podvektor vektora w koji se odnosi na atribute iz skupa G za srpski model, a w'_G za francuski model

Modeli složenije strukture i uključivanje domenskog znanja

- Pretpostavimo da modeliramo prinos pšenice u Srbiji regularizovanom linearnom regresijom:

$$\min_{w} \sum_{i=1}^{N} (y_i - w \cdot x_i)^2 + \lambda \Omega(w)$$

- Uz to, pretpostavimo da imamo na raspolaganju francuski model (koeficijente w' tog modela) i francuski skup podataka
- Skupovi podataka za Francusku i Srbiju ne moraju imati iste atribute
- Neka je G skup atributa koji je zajednički i za Francusku i za Srbiju
- Neka je w_G podvektor vektora w koji se odnosi na atribute iz skupa G za srpski model, a w'_G za francuski model
- Tada ima smisla koristiti sledeći regularizacioni izraz:

$$\Omega(w) = \|w_G - w'_G\|_2^2$$

Modeli složenije strukture i uključivanje domenskog znanja

- Analizirajmo predloženi model:

$$\min_{w} \sum_{i=1}^{N} (y_i - w \cdot x_i)^2 + \lambda \|w_G - w'_G\|_2^2$$

Modeli složenije strukture i uključivanje domenskog znanja

- Analizirajmo predloženi model:

$$\min_{w} \sum_{i=1}^{N} (y_i - w \cdot x_i)^2 + \lambda \| w_G - w'_G \|_2^2$$

- Prilikom rešavanja optimizacionog problema, želimo da imamo što manju grešku nad podacima za Srbiju ali i da norma razlike između srpskog i francuskog modela bude mala

Modeli složenije strukture i uključivanje domenskog znanja

- Analizirajmo predloženi model:

$$\min_{w} \sum_{i=1}^{N} (y_i - w \cdot x_i)^2 + \lambda \|w_G - w'_G\|_2^2$$

- Prilikom rešavanja optimizacionog problema, želimo da imamo što manju grešku nad podacima za Srbiju ali i da norma razlike između srpskog i francuskog modela bude mala
- Regularizacionim metaparametrom λ kontrolišemo šta je važnije od ova birka pri minimizaciji

Modeli složenije strukture i uključivanje domenskog znanja

- Analizirajmo predloženi model:

$$\min_w \sum_{i=1}^{N} (y_i - w \cdot x_i)^2 + \lambda \|w_G - w'_G\|_2^2$$

- Prilikom rešavanja optimizacionog problema, želimo da imamo što manju grešku nad podacima za Srbiju ali i da norma razlike između srpskog i francuskog modela bude mala
- Regularizacionim metaparametrom λ kontrolišemo šta je važnije od ova birka pri minimizaciji
- Ako imamo veliki broj podataka, veći akcenat će biti na minimizaciji greške nad podacima i lambda će biti mala i obrnuto

Modeli složenije strukture i uključivanje domenskog znanja

- Analizirajmo predloženi model:

$$\min_{w} \sum_{i=1}^{N} (y_i - w \cdot x_i)^2 + \lambda \|w_G - w'_G\|_2^2$$

Modeli složenije strukture i uključivanje domenskog znanja

- Analizirajmo predloženi model:

$$\min_{w} \sum_{i=1}^{N} (y_i - w \cdot x_i)^2 + \lambda \|w_G - w'_G\|_2^2$$

- Ukoliko bismo imali više od jednog modela na koji možemo da se oslonimo (na primer, osim za Francusku, poznati su modeli za još nekoliko zemalja), mogli bismo da dodamo po jedan sabirak u regularizacioni izraz za svaki od tih zemalja

Modeli složenije strukture i uključivanje domenskog znanja

- Analizirajmo predloženi model:

$$\min_{w} \sum_{i=1}^{N} (y_i - w \cdot x_i)^2 + \lambda \| w_G - w'_G \|_2^2$$

- Ukoliko bismo imali više od jednog modela na koji možemo da se oslonimo (na primer, osim za Francusku, poznati su modeli za još nekoliko zemalja), mogli bismo da dodamo po jedan sabirak u regularizacioni izraz za svaki od tih zemalja
- Ukoliko ne bismo imali nijedan model na koji možemo da se oslonimo, prethodno znanje bi moglo da se dobije i od domenskog eksperta

Modeli složenije strukture i uključivanje domenskog znanja

- Analizirajmo predloženi model:

$$\min_{w} \sum_{i=1}^{N} (y_i - w \cdot x_i)^2 + \lambda \|w_G - w'_G\|_2^2$$

- Ukoliko bismo imali više od jednog modela na koji možemo da se oslonimo (na primer, osim za Francusku, poznati su modeli za još nekoliko zemalja), mogli bismo da dodamo po jedan sabirak u regularizacioni izraz za svaki od tih zemalja
- Ukoliko ne bismo imali nijedan model na koji možemo da se oslonimo, prethodno znanje bi moglo da se dobije i od domenskog eksperta
- Domenski ekspert bi mogao da kvantifikuje parametre koje doprinose prinosu pšenice i da na taj način zada koeficijent w'_G

Grupna laso regularizacija

- Jedna vrlo korisna vrsta regularizacije je *grupna laso regularizacija* (eng. *group lasso*)

Grupna laso regularizacija

- Jedna vrlo korisna vrsta regularizacije je *grupna laso regularizacija* (eng. *group lasso*)
- Vrlo često, atributi se mogu grupisati prema nekom kriterijumu

Grupna laso regularizacija

- Jedna vrlo korisna vrsta regularizacije je *grupna laso regularizacija* (eng. *group lasso*)
- Vrlo često, atributi se mogu grupisati prema nekom kriterijumu
- U medicinskim primenama, to recimo mogu biti merenja na osnovu uzorka krvi, rendgenskog snimka, snimanja magnetnom rezonancom, biopsije i slično

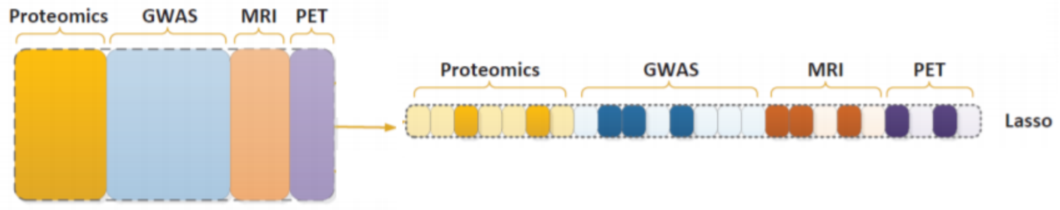
Grupna laso regularizacija

- Jedna vrlo korisna vrsta regularizacije je *grupna laso regularizacija* (eng. *group lasso*)
- Vrlo često, atributi se mogu grupisati prema nekom kriterijumu
- U medicinskim primenama, to recimo mogu biti merenja na osnovu uzorka krvi, rendgenskog snimka, snimanja magnetnom rezonancom, biopsije i slično
- Neke od ovih analiza su neprijatne, neke skupe, na neke se može dugo čekati i slično

Grupna laso regularizacija

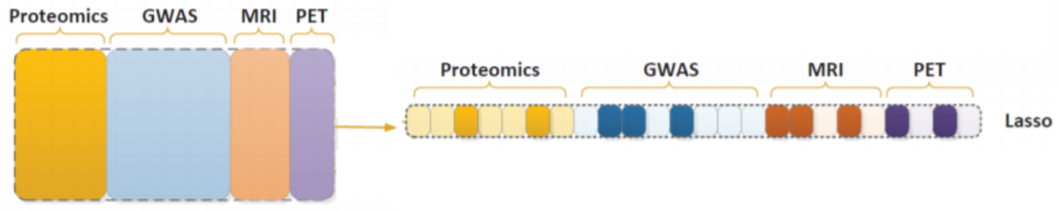
- Jedna vrlo korisna vrsta regularizacije je *grupna laso regularizacija* (eng. *group lasso*)
- Vrlo često, atributi se mogu grupisati prema nekom kriterijumu
- U medicinskim primenama, to recimo mogu biti merenja na osnovu uzorka krvi, rendgenskog snimka, snimanja magnetnom rezonancom, biopsije i slično
- Neke od ovih analiza su neprijatne, neke skupe, na neke se može dugo čekati i slično
- Već je istaknuto da je kvalitet proređenih modela što omogućavaju da se merenja koja odgovaraju određenim atributima ne vrše

Grupna laso regularizacija



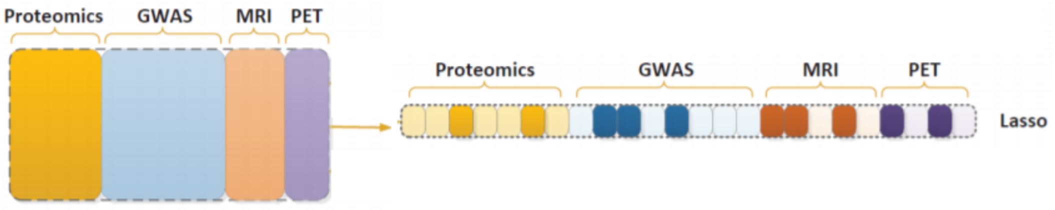
- Ipak, zamislimo da ℓ_1 regularizacija pridruži koeficijent 0 merenju hemoglobina, ali nenula koeficijent merenju triglicerida u krvi

Grupna laso regularizacija



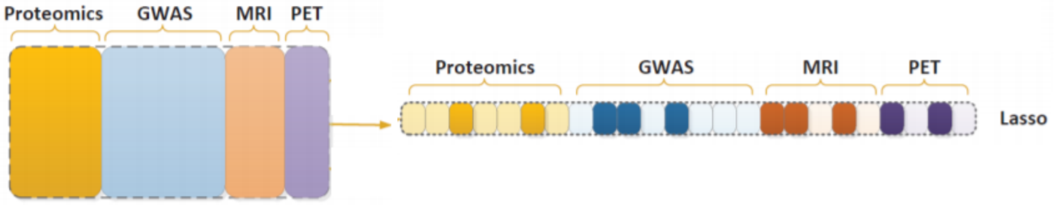
- Ipak, zamislimo da ℓ_1 regularizacija pridruži koeficijent 0 merenju hemoglobina, ali nenula koeficijent merenju triglicerida u krvi
- To što nije potrebno meriti hemoglobin, ipak ne znači mnogo, pošto je svejedno potrebno da pacijent da krv i da se izvrše neke analize

Grupna laso regularizacija



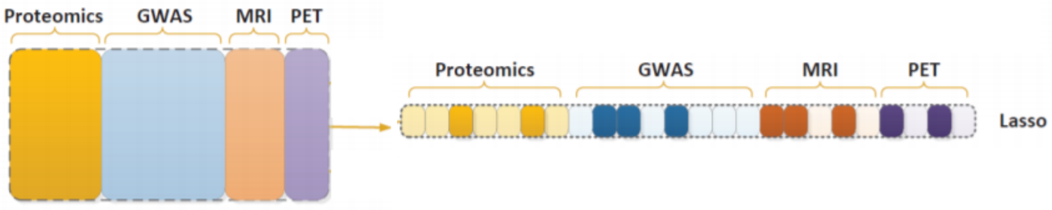
- Ipak, zamislimo da ℓ_1 regularizacija pridruži koeficijent 0 merenju hemoglobina, ali nenula koeficijent merenju triglicerida u krvi
- To što nije potrebno meriti hemoglobin, ipak ne znači mnogo, pošto je svejedno potrebno da pacijent da krv i da se izvrše neke analize
- Suštinski dobitak bi bio da nijednu od analiza krvi nije potrebno raditi

Grupna laso regularizacija



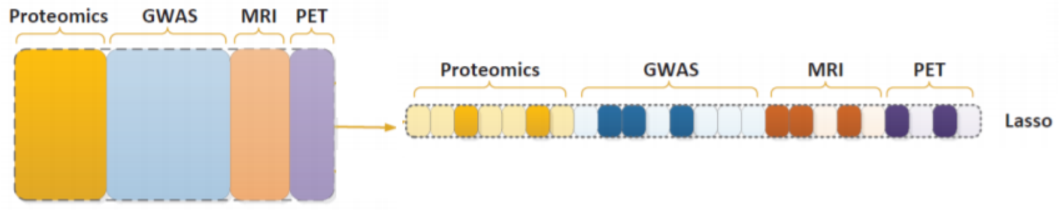
- Ipak, zamislimo da ℓ_1 regularizacija pridruži koeficijent 0 merenju hemoglobina, ali nenula koeficijent merenju triglicerida u krvi
- To što nije potrebno meriti hemoglobin, ipak ne znači mnogo, pošto je svejedno potrebno da pacijent da krv i da se izvrše neke analize
- Suštinski dobitak bi bio da nijednu od analiza krvi nije potrebno raditi
- U suprotnom, u ovom slučaju, mogu se uraditi sve

Grupna laso regularizacija



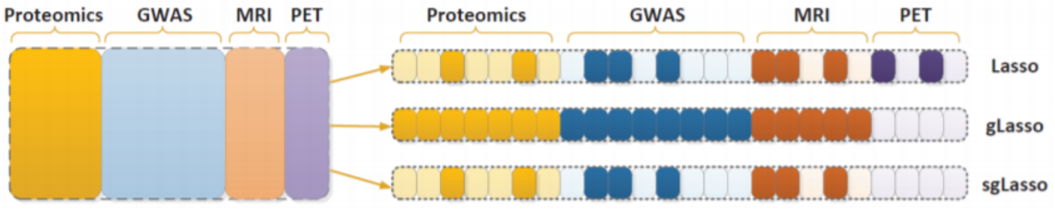
- Ipak, zamislimo da ℓ_1 regularizacija pridruži koeficijent 0 merenju hemoglobina, ali nenula koeficijent merenju triglicerida u krvi
- To što nije potrebno meriti hemoglobin, ipak ne znači mnogo, pošto je svejedno potrebno da pacijent da krv i da se izvrše neke analize
- Suštinski dobitak bi bio da nijednu od analiza krvi nije potrebno raditi
- U suprotnom, u ovom slučaju, mogu se uraditi sve
- Slično, ukoliko se uradi snimanje magnetnom rezonancom, nebitno je da li se računaju neki atributi snimka ili svi

Grupna laso regularizacija



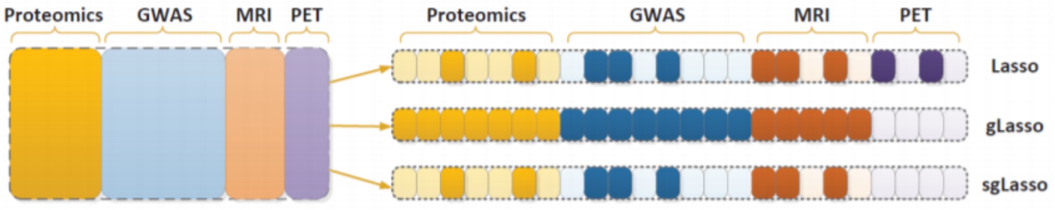
- Ipak, zamislimo da ℓ_1 regularizacija pridruži koeficijent 0 merenju hemoglobina, ali nenula koeficijent merenju triglicerida u krvi
- To što nije potrebno meriti hemoglobin, ipak ne znači mnogo, pošto je svejedno potrebno da pacijent da krv i da se izvrše neke analize
- Suštinski dobitak bi bio da nijednu od analiza krvi nije potrebno raditi
- U suprotnom, u ovom slučaju, mogu se uraditi sve
- Slično, ukoliko se uradi snimanje magnetnom rezonancom, nebitno je da li se računaju neki atributi snimka ili svi
- Grupna laso regularizacija rešava ovaj problem.

Grupna laso regularizacija



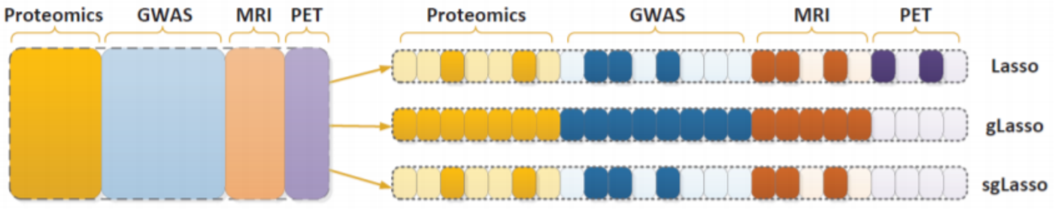
- Kako postižemo proređenost po grupama?

Grupna laso regularizacija



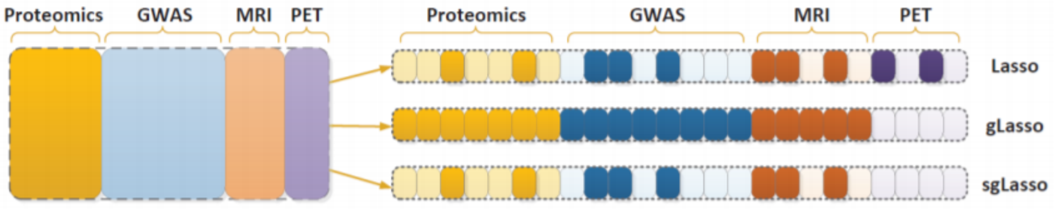
- Kako postižemo proređenost po grupama?
- Ako u regularizacioni izraz stavimo normu nekog vektora, tada taj vektor ide u nulu

Grupna laso regularizacija



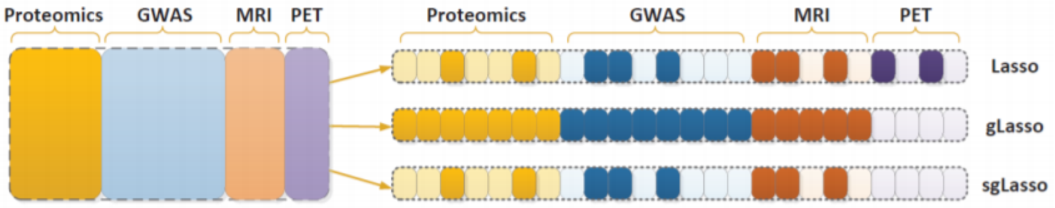
- Kako postižemo proređenost po grupama?
- Ako u regularizacioni izraz stavimo normu nekog vektora, tada taj vektor ide u nulu
- Potrebno je formirati grupe atributa $\{G_1, \ldots, G_M\}$ koje mogu, ali ne moraju biti disjunktne

Grupna laso regularizacija



- Kako postižemo proređenost po grupama?
- Ako u regularizacioni izraz stavimo normu nekog vektora, tada taj vektor ide u nulu
- Potrebno je formirati grupe atributa $\{G_1, \ldots, G_M\}$ koje mogu, ali ne moraju biti disjunktne
- w_{G_i} predstavljaju koeficijente modela koji odgovaraju atributima grupe w_{G_i}

Grupna laso regularizacija



- Kako postižemo proređenost po grupama?
- Ako u regularizacioni izraz stavimo normu nekog vektora, tada taj vektor ide u nulu
- Potrebno je formirati grupe atributa $\{G_1, \ldots, G_M\}$ koje mogu, ali ne moraju biti disjunktne
- w_{G_i} predstavljaju koeficijente modela koji odgovaraju atributima grupe w_{G_i}
- Grupna laso regularizacija se vrši upotrebom regularizacionog izraza

$$\Omega(w) = \sum_{i=1}^{M} \|w_{G_i}\|_2$$

Grupna laso regularizacija

- Grupna laso regularizacija se vrši upotrebom regularizacionog izraza

$$\Omega(w) = \sum_{i=1}^{M} \|w_{G_i}\|_2$$

Grupna laso regularizacija

- Grupna laso regularizacija se vrši upotrebom regularizacionog izraza

$$\Omega(w) = \sum_{i=1}^{M} \|w_{G_i}\|_2$$

- Primetimo da norma nije kvadrirana, jer da jeste, to bi bila obična ℓ_2 norma svih koeficijenata, ne bi se pojedinačno fokusirala na određene grupe

Grupna laso regularizacija

- Grupna laso regularizacija se vrši upotrebom regularizacionog izraza

$$\Omega(w) = \sum_{i=1}^{M} \|w_{G_i}\|_2$$

- Primetimo da norma nije kvadrirana, jer da jeste, to bi bila obična ℓ_2 norma svih koeficijenata, ne bi se pojedinačno fokusirala na određene grupe
- Ovakva regularizacija teži tome da anulira norme pojedinačnih grupa, a anuliranjem normi grupa anuliraju se svi koeficijenti u grupi

Grupna laso regularizacija

- Grupna laso regularizacija se vrši upotrebom regularizacionog izraza

$$\Omega(w) = \sum_{i=1}^{M} \|w_{G_i}\|_2$$

- Primetimo da norma nije kvadrirana, jer da jeste, to bi bila obična ℓ_2 norma svih koeficijenata, ne bi se pojedinačno fokusirala na određene grupe
- Ovakva regularizacija teži tome da anulira norme pojedinačnih grupa, a anuliranjem normi grupa anuliraju se svi koeficijenti u grupi
- Pored anuliranja celih grupa, i dalje može biti poželjno da se i u okviru relevantnih grupa model što više proredi

Grupna laso regularizacija

- Grupna laso regularizacija se vrši upotrebom regularizacionog izraza

$$\Omega(w) = \sum_{i=1}^{M} \|w_{G_i}\|_2$$

- Primetimo da norma nije kvadrirana, jer da jeste, to bi bila obična ℓ_2 norma svih koeficijenata, ne bi se pojedinačno fokusirala na određene grupe
- Ovakva regularizacija teži tome da anulira norme pojedinačnih grupa, a anuliranjem normi grupa anuliraju se svi koeficijenti u grupi
- Pored anuliranja celih grupa, i dalje može biti poželjno da se i u okviru relevantnih grupa model što više proredi
- U tom slučaju moguće je koristiti kombinaciju grupne i obične laso regularizacije:

$$\Omega(w) = \mu\|w\|_1 + (1-\mu)\sum_{i=1}^{M} \|w_{G_i}\|_2$$

za neko $\mu \in [0,1]$

Grupna laso regularizacija

- Još jedan kontekst u kojem je grupna laso regularizacija vrlo korisna je postojanje kategoričkih atributa.

Grupna laso regularizacija

- Još jedan kontekst u kojem je grupna laso regularizacija vrlo korisna je postojanje kategoričkih atributa.
- Kao što je već rečeno kategorički atribut se najčešće uključuje u model tako što se kodira pomoću više binarnih promenljivih.

Grupna laso regularizacija

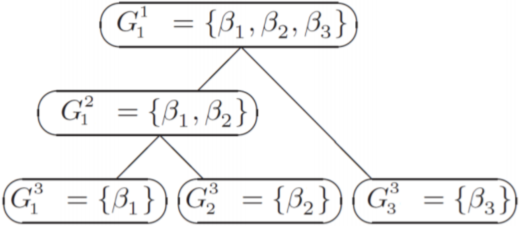
- Još jedan kontekst u kojem je grupna laso regularizacija vrlo korisna je postojanje kategoričkih atributa.
- Kao što je već rečeno kategorički atribut se najčešće uključuje u model tako što se kodira pomoću više binarnih promenljivih.
- Isključivanje samo nekih od tih binarnih promenljivih neće omogućiti potpuno izbacivanje atributa iz modela i eliminisati potrebu za njegovim merenjem.

Grupna laso regularizacija

- Još jedan kontekst u kojem je grupna laso regularizacija vrlo korisna je postojanje kategoričkih atributa.
- Kao što je već rečeno kategorički atribut se najčešće uključuje u model tako što se kodira pomoću više binarnih promenljivih.
- Isključivanje samo nekih od tih binarnih promenljivih neće omogućiti potpuno izbacivanje atributa iz modela i eliminisati potrebu za njegovim merenjem.
- S druge strane, korišćenje grupne laso regularizacije, vodiće tome da su sve binarne promenljive uključene u model ili da su sve isključene iz njega.

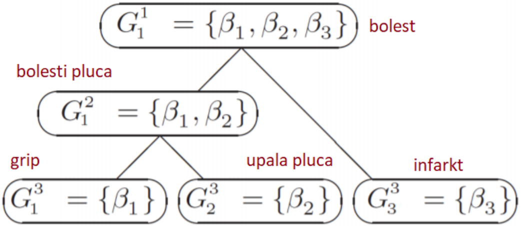
Grupna laso regularizacija

- Još jedan vid regularizacije koja daje specifičnu strukturu modela, a i uključivanje domenskog znanja je *hijerarhijska laso regularizacija* (eng. *tree group lasso*), pri čemu pretpostavljamo da je hijerarhija data stablom u čijim su listovima (i samo u listovima) atributi



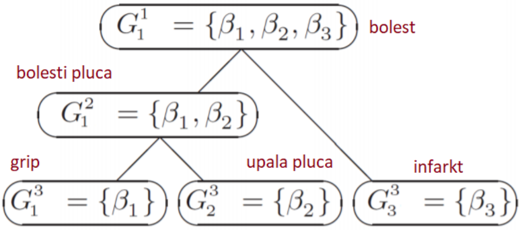
Grupna laso regularizacija

- Na primer, medicinske dijagnoze se prirodno organizuju u stabla, pri čemu u korenu stabla može biti najopštija dijagnoza *bolest*, na nešto nižim nivoima mogu biti recimo grupe dijagnoza *bolest pluća*, *bolest srca* i tako dalje, dok u listovima mogu biti konkretne dijagnoze poput *grip*, *upala pluća*, *artritis* i tako dalje



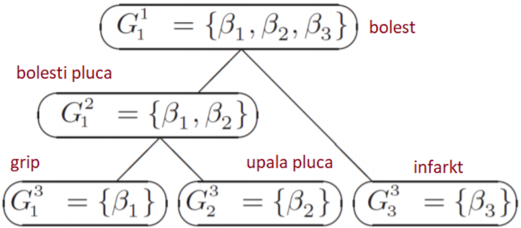
Grupna laso regularizacija

- U ovom slučaju regularizacioni izraz ima istu formu kao i izraz za grupnu laso regularizaciju



Grupna laso regularizacija

- U ovom slučaju regularizacioni izraz ima istu formu kao i izraz za grupnu laso regularizaciju
- Hijerarhijski efekat se postiže tako što svakom čvoru stabla odgovara grupa svih atributa koji su u listovima potomcima datog čvora



Grupna laso regularizacija

- U ovom slučaju regularizacioni izraz ima istu formu kao i izraz za grupnu laso regularizaciju
- Hijerarhijski efekat se postiže tako što svakom čvoru stabla odgovara grupa svih atributa koji su u listovima potomcima datog čvora
- Na taj način, ukoliko se regularizacijom iz modela eliminišu plućne bolesti, automatski su eliminisane i sve pojedinačne plućne bolesti.