

Polu-nadzirano učenje

Jurica Levatić

Pregled

1) Uvod u polu-nadzirano učenje

2) Klase algoritama

- Generativni modeli
- Low-density Separation
- Graph-based metode

3) Zaključak

Zašto?

Ljudi žele bolje rezultate besplatno.

Nadzirano i ne-nadzirano učenje

- Nadzirano učenje
 - učenje na temelju unaprijed labeliranih podataka
- Polu-nadzirano učenje
 - koristi labelirane i ne-labelirane podatke za izgradnju (boljih) klasifikatora
- Nenadzirano učenje
 - nalaženje korisnih struktura na podacima

Motivacija

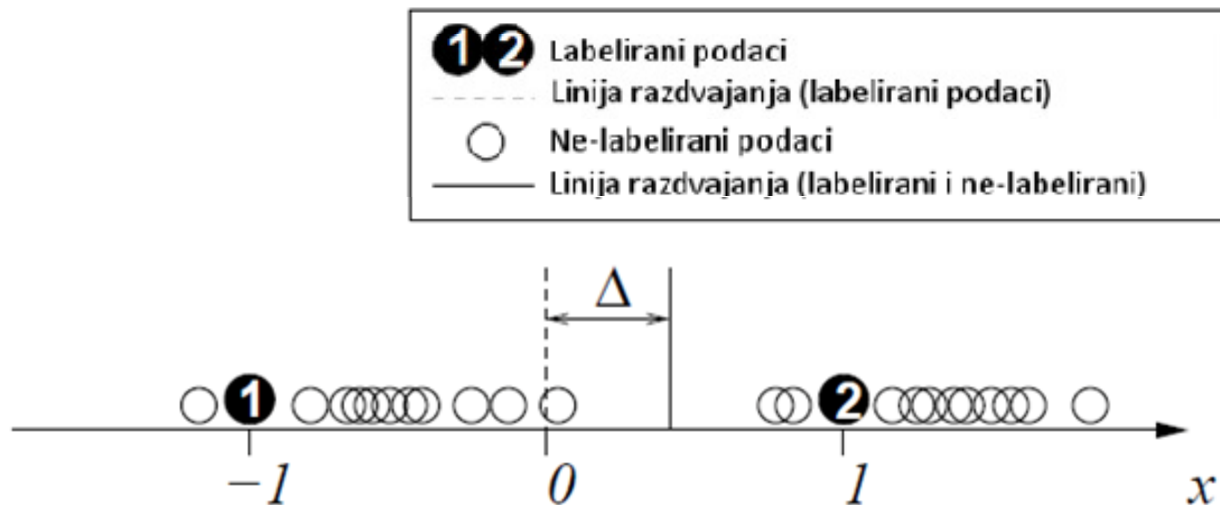
- ne-labelirani podaci su jeftini
- labelirani podaci mogu biti skupi i teško dostupni
 - anotiranje telefonskog razgovora
 - klasificiranje web stranica

Notacije

- ulazni podaci x , labele y
- labelirani
podaci $(X_l, Y_l) = \{(x_1, y_1), \dots, (x_l, y_l)\}$
- ne-labelirani podaci $X_u = \{x_{l+1}, \dots, x_{l+u}\}$
- učimo $f : X \rightarrow Y$
- obično $l \ll u$
- test podaci $X_{test} = \{x_{l+u+1}, \dots\}$

Mogu li ne-labelirani podaci pomoći?

Mogu, uz određene pretpostavke.



Self Training Algoritam

Ulaz: labelirani podaci (X_l, Y_l) , ne-labelirani podaci X_u

- 1) Treniraj f na (X_l, Y_l)
- 2) Napravi predikcije na $x \in X_u$
- 3) Dodaj $(x, f(x))$ u (X_l, Y_l)
- 4) Ponavljaj

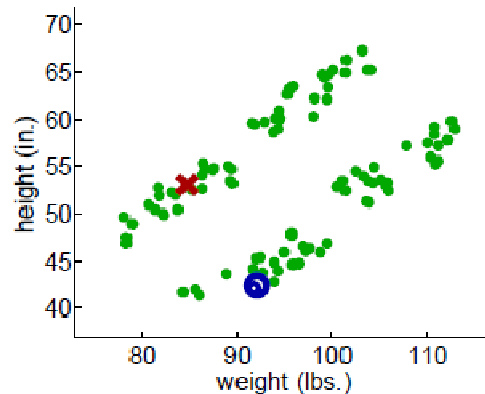
Pretpostavka: Pouzdane predikcije su točne.

Self Training Algoritam

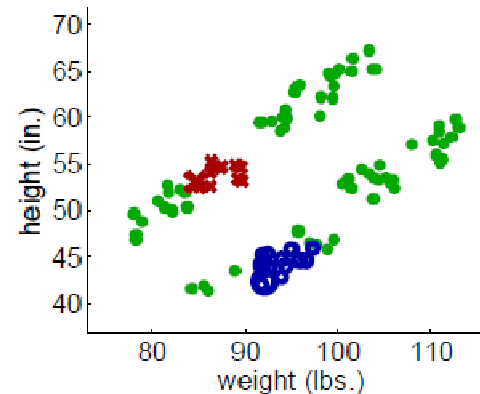
Varijacije (korak 3):

- dodaj nekoliko najvjerojatnijih $(x, f(x))$ labeliranim podacima
- dodaj sve $(x, f(x))$ labeliranim podacima
- dodaj sve $(x, f(x))$ labeliranim podacima, sa težinama prema vjerojatnosti predikcija
- ...

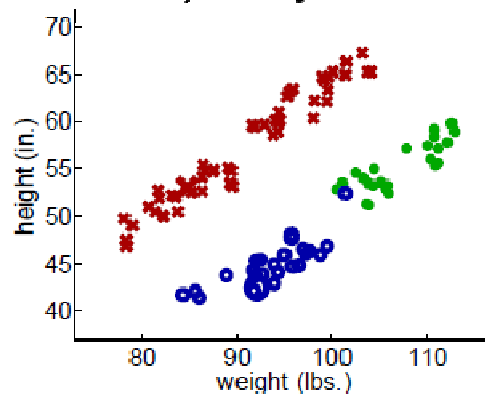
Self Training primjer: $f = 1\text{NN}$



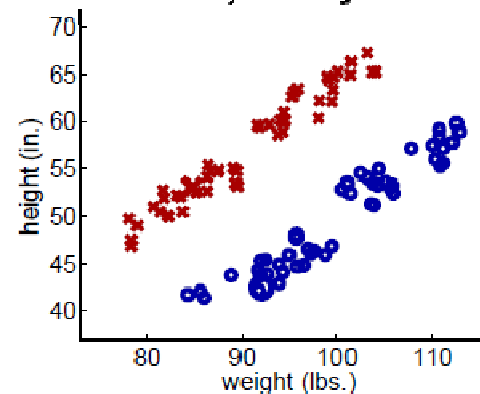
a) iteracija 1



b) iteracija 25

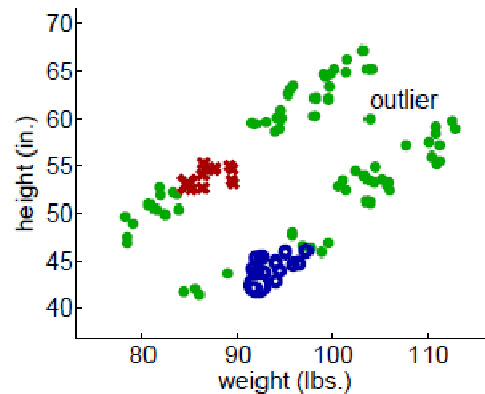


c) iteracija 74

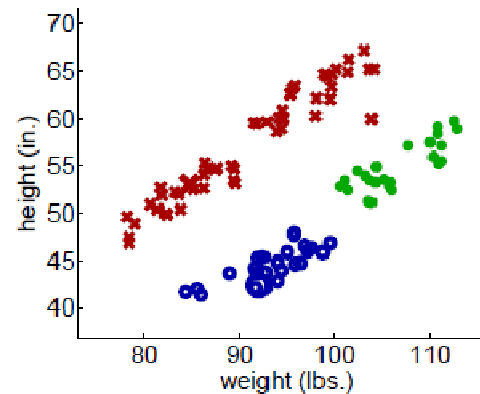


d) finalno labeliranje

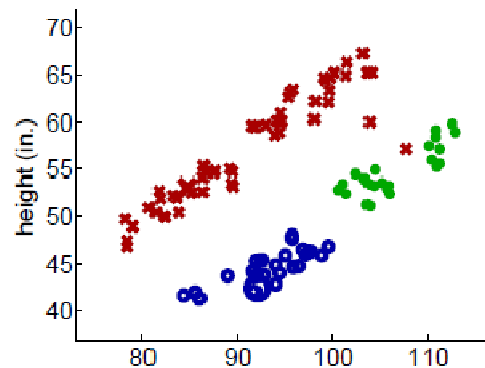
Self Training primjer: $f = 1\text{NN}$



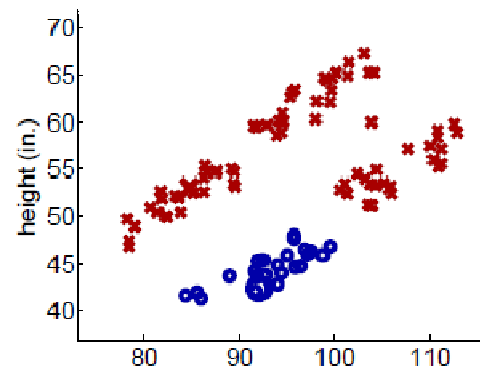
a)



b)



c)



d)

Pregled

1) Uvod u polu-nadzirano učenje

2) Klase algoritama

- Generativni modeli
- Low-density Separation
- Graph-based metode

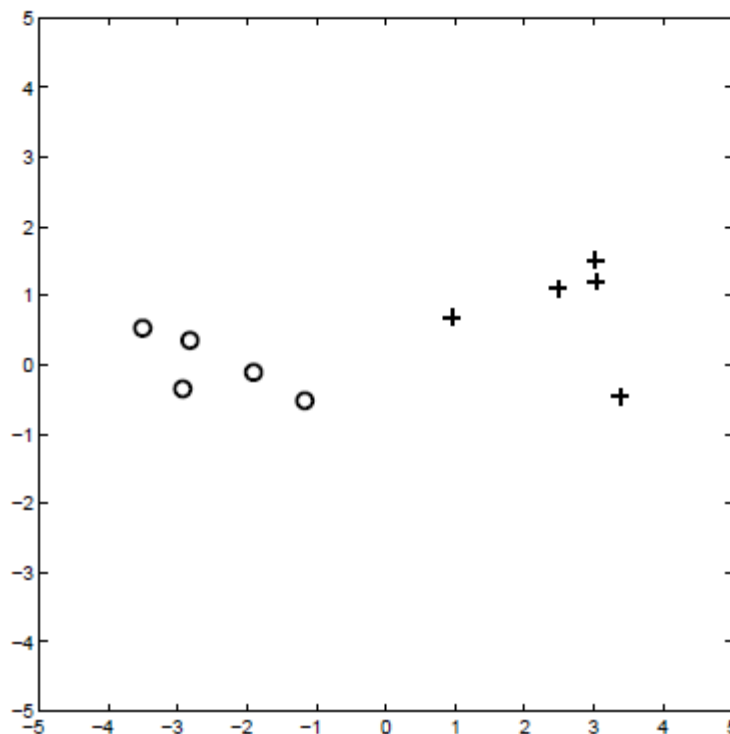
3) Zaključak

Općenito

- pretpostavka generativnog modela (gustoće)
 $p(x, y) = p(y)p(x|y)$, gdje je $p(x|y)$ mješavina razdioba
- pomoću (velike količine) ne-labeliranih podataka mogu se odrediti komponente razdiobe (najvjerojatniji parametri!)
- ako je **pretpostavka modela dobra**, ne-labelirani podaci **sigurno** poboljšavaju točnost (Castelli & Cover, 1995/96., Ratsby & Venkatesh 1995.)

Primjer

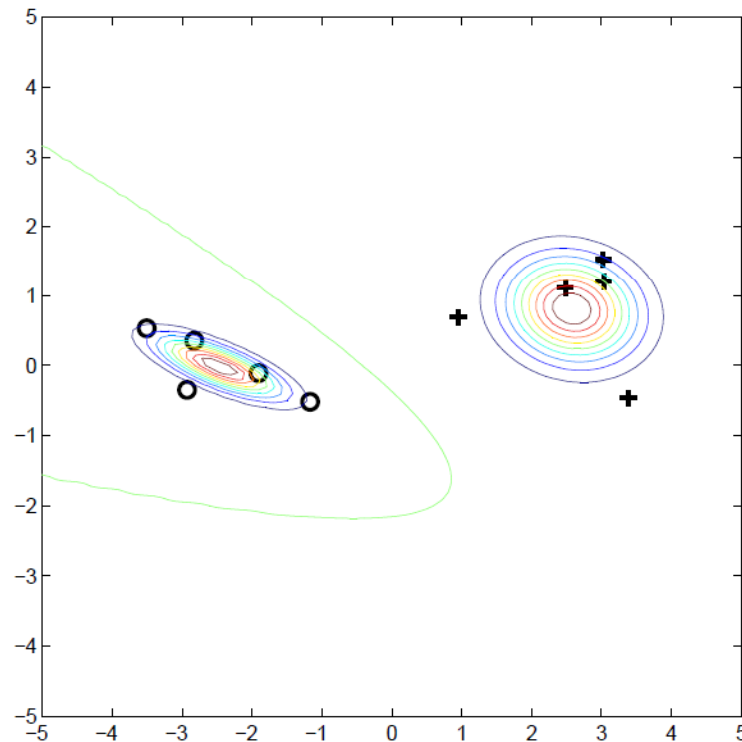
Labelirani podaci (X_l, Y_l) :



Uz pretpostavku da su obje klase normalno distribuirane, gdje leži *decision boundary*?

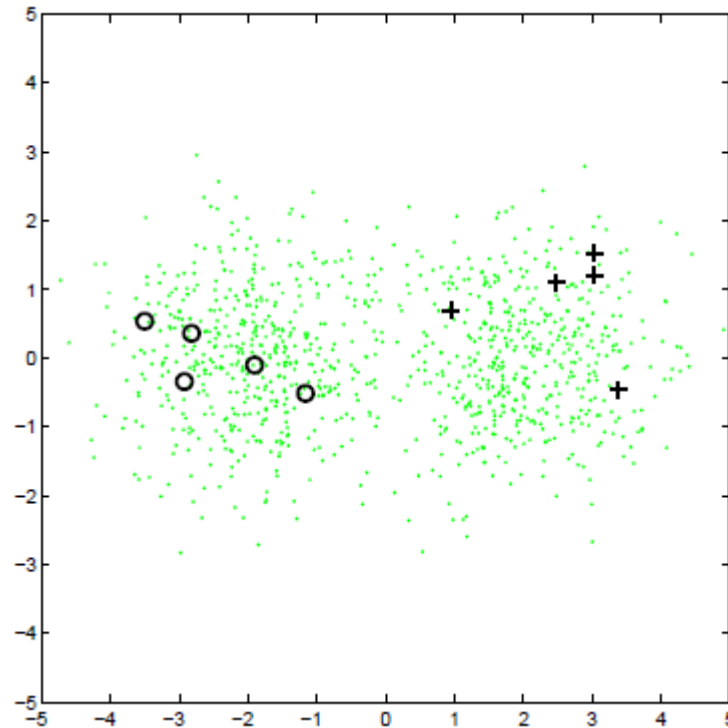
Primjer

Najvjerojatniji model:



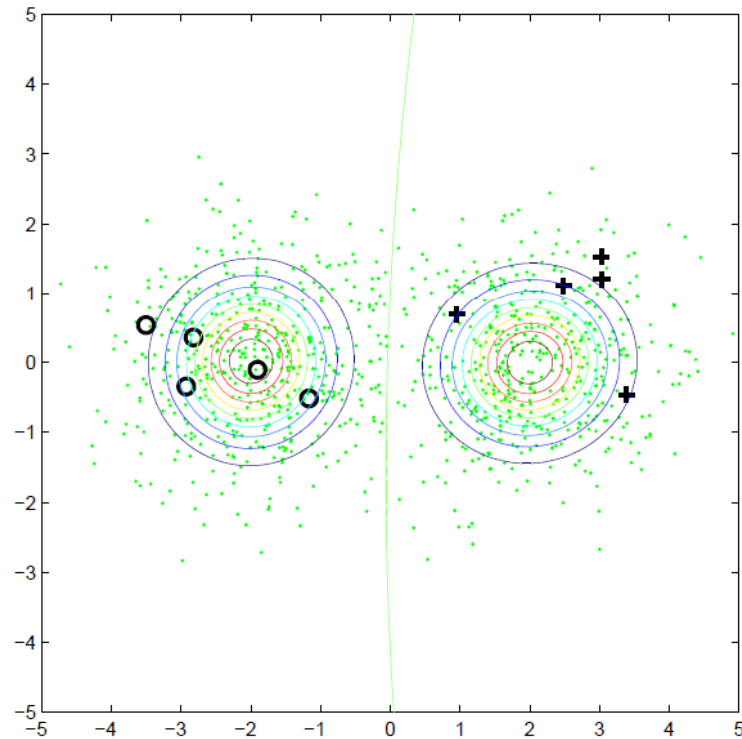
Primjer

Dodajemo ne-labelirane podatke:



Primjer

Najvjerojatniji model:



Primjer 2 (binarni klasifikacijski problem)

Generativni model: mješavina Gaussovih distribucija

1) Labelirani podaci

- $\log p(X_l, Y_l | \theta) = \sum_{i=0}^l \log p(y_i | \theta) p(x_i | y_i, \theta)$
- nalazimo MLE procjenitelj za θ (parametri modela)

2) Labelirani i ne-labelirani podaci

- $\log p(X_l, Y_l, X_u | \theta) = \sum_{i=0}^l \log p(y_i | \theta) p(x_i | y_i, \theta) + \sum_{i=l+1}^{l+u} \log \left(\sum_{y=1}^2 p(y_i | \theta) p(x_i | y_i, \theta) \right)$
- pomoću EM algoritma nalazimo (lokalno) najvjerojatniji θ

Primjer 2 (binarni klasifikacijski problem)

EM algoritam:

1) početni: $\theta = \text{MLE za } (X_l, Y_l)$

2) Ponavljaj:

- (E-korak) za sve $x \in X_u$ izračunaj očekivane labele
$$p(y|x, \theta) = \frac{p(y|\theta)p(x|y, \theta)}{p(x|\theta)}$$
- (M-korak) izračunaj MLE θ sa (labeliranim) X_u

Često korišteni modeli

- Mješavina Gaussovih distribucija
 - klasifikacija slika
 - EM algoritam
- Mješavina multinomialnih razdioba
 - kategorizacija teksta
 - EM algoritam
- Skriveni Markovljevi modeli (HMM)
 - prepoznavanje govora
 - Baum – Welch algoritam

Sažetak

- **cilj:** maksimizirati $p(X_l, Y_l, X_u | \theta)$
- mogu biti jako učinkoviti ako je pretpostavka modela blizu točnog
- polu-nadzirano učenje se jače oslanja na pretpostavku točnosti modela
- često teško provjeriti točnost modela
- ne-labelirani podaci mogu smanjiti točnost ako je model pogrešan

Pregled

1) Uvod u polu-nadzirano učenje

2) Klase algoritama

- Generativni modeli
- **Low-density Separation**
- Graph-based metode

3) Zaključak

Low-density Separation metode

Pretpostavka:

Linija razdvajanja leži u području male gustoće.

Ekvivalentno(*Cluster Assumption*):

Ako su točke u istom clusteru vjerojatno su i u istoj klasi.

Semi-supervised SVM (S^3VM)

- Transductive SVM (TSVM)
- nadogradnja standardnog SVM

SVM ponavljanje

Pronaći **hiperravninu razdvajanja**: $w^\tau x + b = 0$
takvu da je **margina** ρ maksimalna.

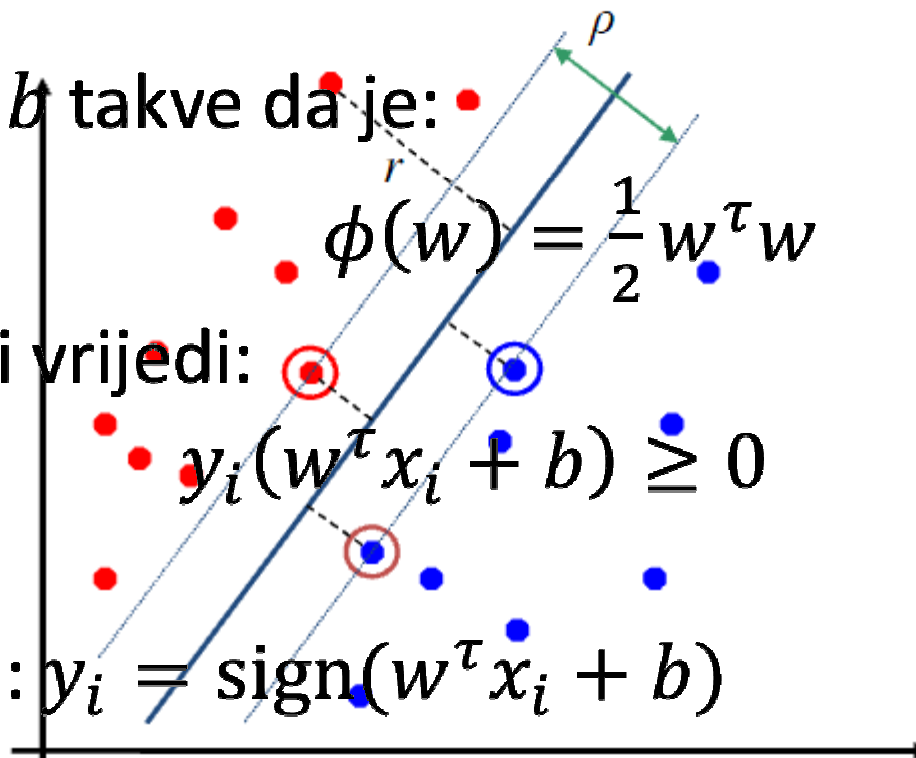
Pronaći w i b takve da je:

$$\phi(w) = \frac{1}{2} w^\tau w$$

minimalno i vrijedi:

$$y_i (w^\tau x_i + b) \geq 0$$

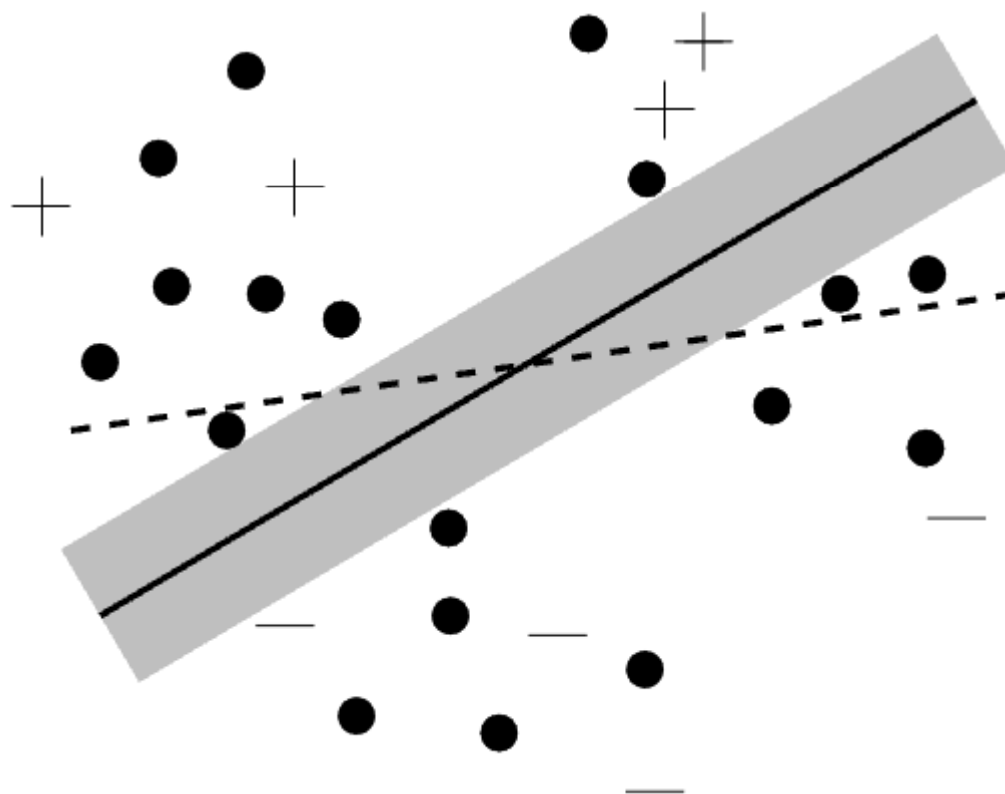
Klasifikacija: $y_i = \text{sign}(w^\tau x_i + b)$



S^3VM

Maksimizirati marginu na labeliranim i ne-labeliranim podacima.

--- SVM
— S^3VM



S³VM

- minimizirati: $\phi(y_{u_1}^*, \dots, y_{u_u}^*, w, b) = \frac{1}{2} w^\tau w$ uz

$$\forall_{i=1}^l: y_i(w x_i + b) \geq 0,$$

$$\forall_{j=l+1}^u y_j^*(w x_j + b) \geq 0$$
- pronaći labeliranje $y_{u_1}^*, \dots, y_{u_u}^*$ takvo da je margina na (labeliranim i ne-labeliranim) podacima maksimalna
- napraviti jedan standardni SVM za svako labeliranje od X_u i odabrati onaj sa najvećom marginom

$$\text{margina } \rho : \min_{x \in X_l \cup X_u} \left[\frac{y_i}{\|w\|} (w^\tau x + b) \right]$$

S³VM prednosti/nedostaci

- posebno dobra metoda za probleme klasifikacije teksta
- često dobra u slučajevima: mali *train test* - veliki *test set*
 - problem: težak optimizacijski problem
 - postojeći algoritmi → lokalni optimum

Pregled

1) Uvod u polu-nadzirano učenje

2) Klase algoritama

- Generativni modeli
- Low-density Separation
- Graph-based metode

3) Zaključak

Općenito

Podaci: $X_l \cup X_u \rightarrow$ graf

- vrhovi grafa - podaci (labelirani i ne-labelirani)
- lukovi - sličnost vrhova

Sličnost (npr.):

$$W_{i,j} = \begin{cases} 1, & \text{ako je } x_i \text{ među } k \text{ najbilžih susjeda od } x_j \\ 0, & \text{inače} \end{cases}$$

$$W_{i,j} = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$$

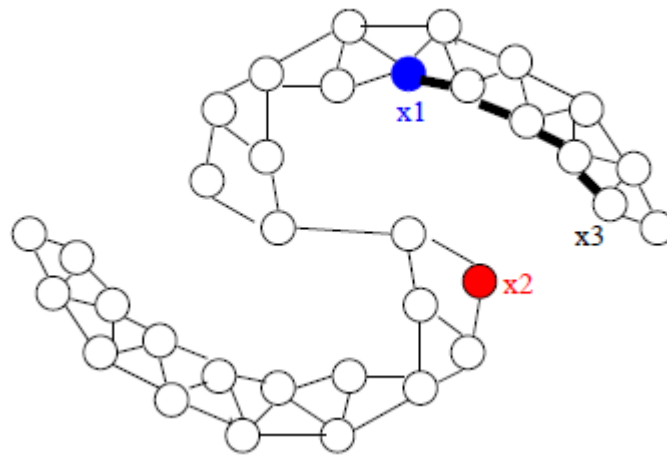
općenito: $W_{i,j}$ = simetrična pozitivna funkcija

Graph-based metode

Pretpostavka: Primjeri povezani lukom velike težine vjerojatno pripadaju istoj klasi.

Graph-based metode

Ideja: Propagirati poznate labele kroz graf da bi labelirali sve vrhove.



Mincut algoritam

Za fiksno Y_l pronači $Y_u \in \{0,1\}^{n-1}$ tako da je

$$\sum_{i,j} w_{i,j} |y_i - y_j|$$

minimalno.

- Kombinatorni problem – rješenje u polinomijalnom vremenu

Sažetak

1) Konstruirati (težinski) graf iz $X_l \cup X_u$

- najvažnij korak?
- nedovoljno istraženo

2) Propagirati labele

- Laplacian regularization
- Gaussian fields and harmonic function
- ...

Zaključak

- labelirani podaci su skupi i ima ih malo
- ne-labelirani jeftini i ima ih puno
- mogu pomoći uz dobre pretpostavke
- labelirani \ll ne-labelirani \rightarrow performanse
- polu-nadzirano učenje je još u stanju „ranog djetinjstva”

Literatura

- Olivier Chapelle, Alexander Zien, Bernhard Schölkopf (Eds.), (2006), *Semi-supervised learning*, MIT Press.
- Xiaojin Zhu (2005), *Semi-supervised learning literature survey*, TR-1530, University of Wisconsin-Madison Department of Computer Science.
- Xiaojin Zhu (2007), *Semi-Supervised Learning Tutorial*, University of Wisconsin-Madison Department of Computer Science.
- Jason Weston, *Large-Scale Semi-Supervised Learning*, NEC Laboratories America, Inc.
- Tomislav Šmuc, predavanja iz kolegija Strojno učenje, 2010.