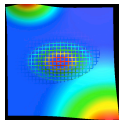
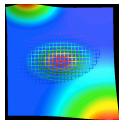


Nagy adathalmazok labor

2018-2019 őszi félév

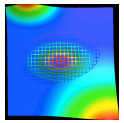
2018.10.10

1. Maximális margó
2. Generalizáltság bevezetés



Féléves terv

- Kiértékelés: cross-validation, bias-variance trade-off
- Supervised learning (discr. és generative modellek): nearest neighbour, decision tree (döntési fák), logisztikus regresszió, nem lineáris osztályozás, neurális hálózatok, support vector machine, idős osztályozás és dynamic time warping
- Lineáris, polinomiális és sokdimenziós regresszió és optimalizálás
- Tree ensembles: AdaBoost, GBM, Random Forest
- Clustering (klaszterezés): k-means, sűrűség és hierarchikus modellek
- Principal component analysis (PCA, SVD), collaborative filtering, ajánlórendszerek
- Anomália keresés
- Gyakori termékhalmazok
- Szublineáris algoritmusok: streaming, Johnson-Lindenstrauss, count-min sketch



Következő hetek

Október 3: NB, logisztikus regresszió

Október 10: generalizáció polinomiális szeparálás, SVM

Október 11: LR és SVM ipython

Október 17: ANN, MLP

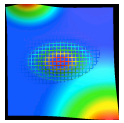
Október 24: CNN, RBM

Október 25: RNN, LSTM

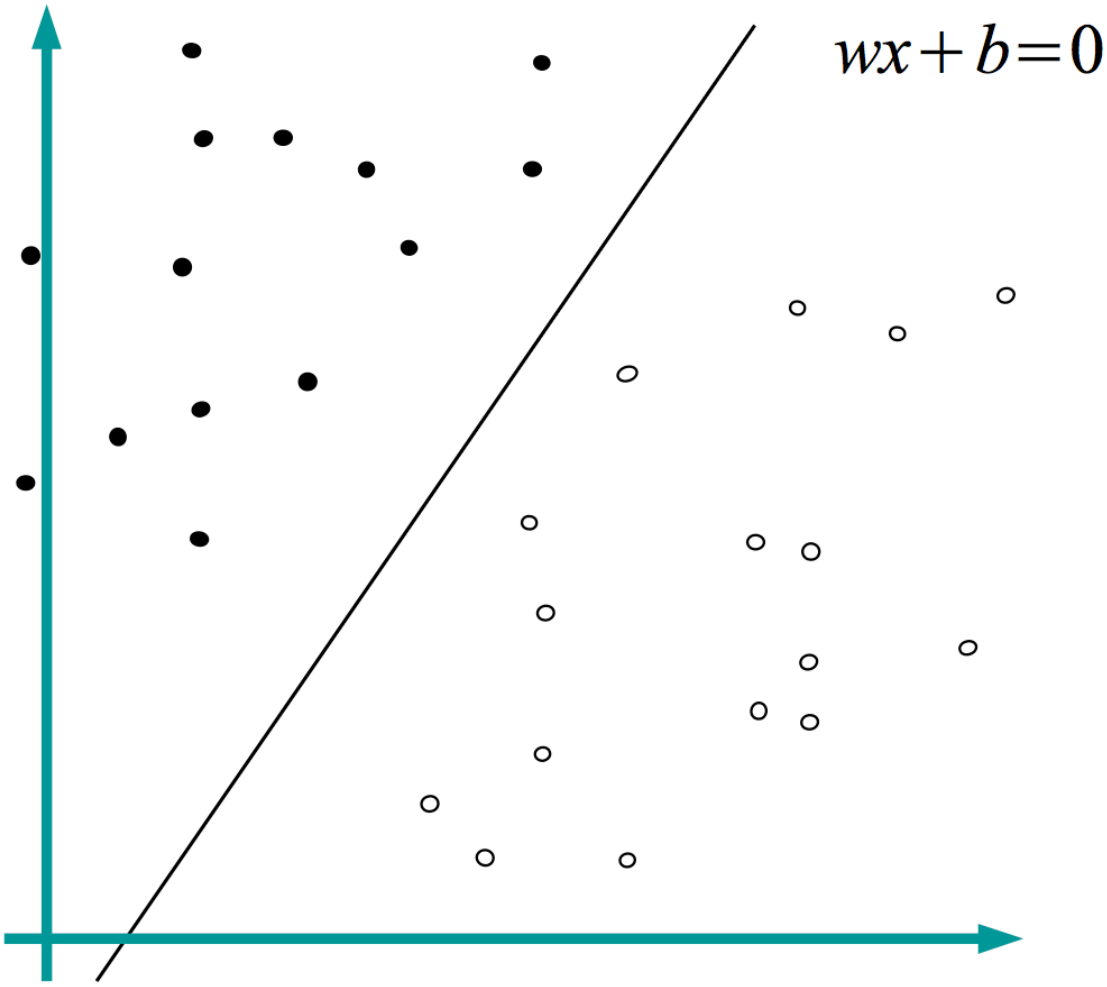
Október 31: nem lesz ora

November 6: VAE, GAN, DBN

November 7: TensorFlow, Keras ipython



Lineáris szeparálás



Logisztikus regresszió

Jobbra balra: Legyen p(x) az első osztályra döntés valószínűsége és tegyük fel, hogy:

$$\ln \frac{p(x)}{1-p(x)} \approx x^T \omega + \omega_0$$

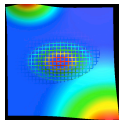
Melyből: $$p(x \mid \omega) = sigm(x \mid \omega) = \frac{1}{1+\mathrm{e}^{-(x^T\omega+\omega_0)}}$$

Amennyiben a log-likelihood-ra optimalizálunk (tanulóhalmaz $X=\{x_1,..,x_t\}$) és a mintáinkat függetlennek tekintjük:

$$\frac{\partial \mathcal{L}(\omega \mid X)}{\partial \omega_i} = \sum_{x_t \in X^{(+)}} \frac{\partial \ln p(x_t \mid \omega)}{\partial \omega_i} + \sum_{x_t \in X^{(-)}} \frac{\partial \ln(1-p(x_t \mid \omega))}{\partial \omega_i}$$

$$= \sum_{x_t \in X^{(+)}} (1-p(x_t \mid \omega))x_{ti} - \sum_{x_t \in X^{(-)}} p(x_t \mid \omega)x_{ti}$$

$$= \sum_{x_t \in \{X^{(-)},X^{(+)}\}} (y_t - p(x_t \mid \omega))x_{ti}$$



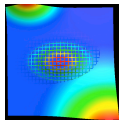
Lineáris szeparálhatóság

Lehetséges-e és ha igen, hogyan a következő függvényeket megadni lineáris szeparátorral?

- OR
- AND

Keressük meg gradiens módszerrel a következő függvény maximumát:

sin(x)



Lineáris szeparálhatóság

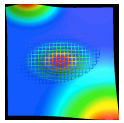
Lehetséges-e és ha igen, hogyan a következő függvényeket megadni lineáris szeparátorral?

- OR
- AND

Keressük meg gradiens módszerrel a következő függvény maximumát:

sin(x)

Majd pedig a következőét: x*sin(x)



Lineáris szeparálhatóság

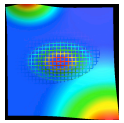
Lehetséges-e és ha igen, hogyan a következő függvényeket megadni lineáris szeparátorral?

- OR
- AND

Keressük meg gradiens módszerrel a következő függvény maximumát:

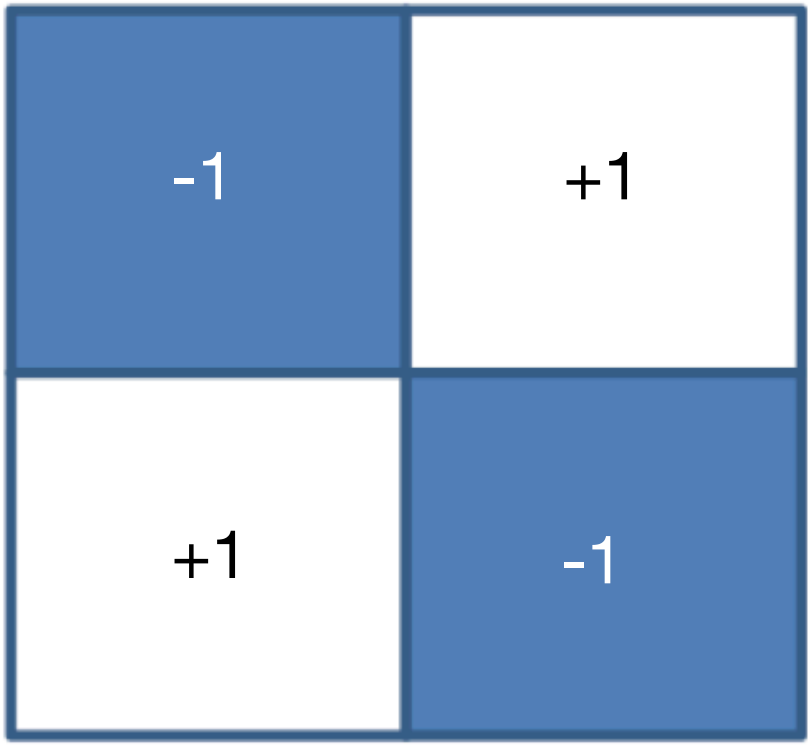
sin(x)

Majd pedig a következőét: x*sin(x)
Ha x>-10 és x<10…

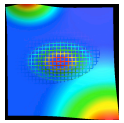


Lineáris szeparálhatóság

Lehetséges-e lineáris szeparálót találni a lenti adathalmazhoz?



Mit is jelent ez számunkra a lineáris osztályozók szempontjából? De mielőtt ...

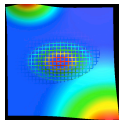


Maximális margó

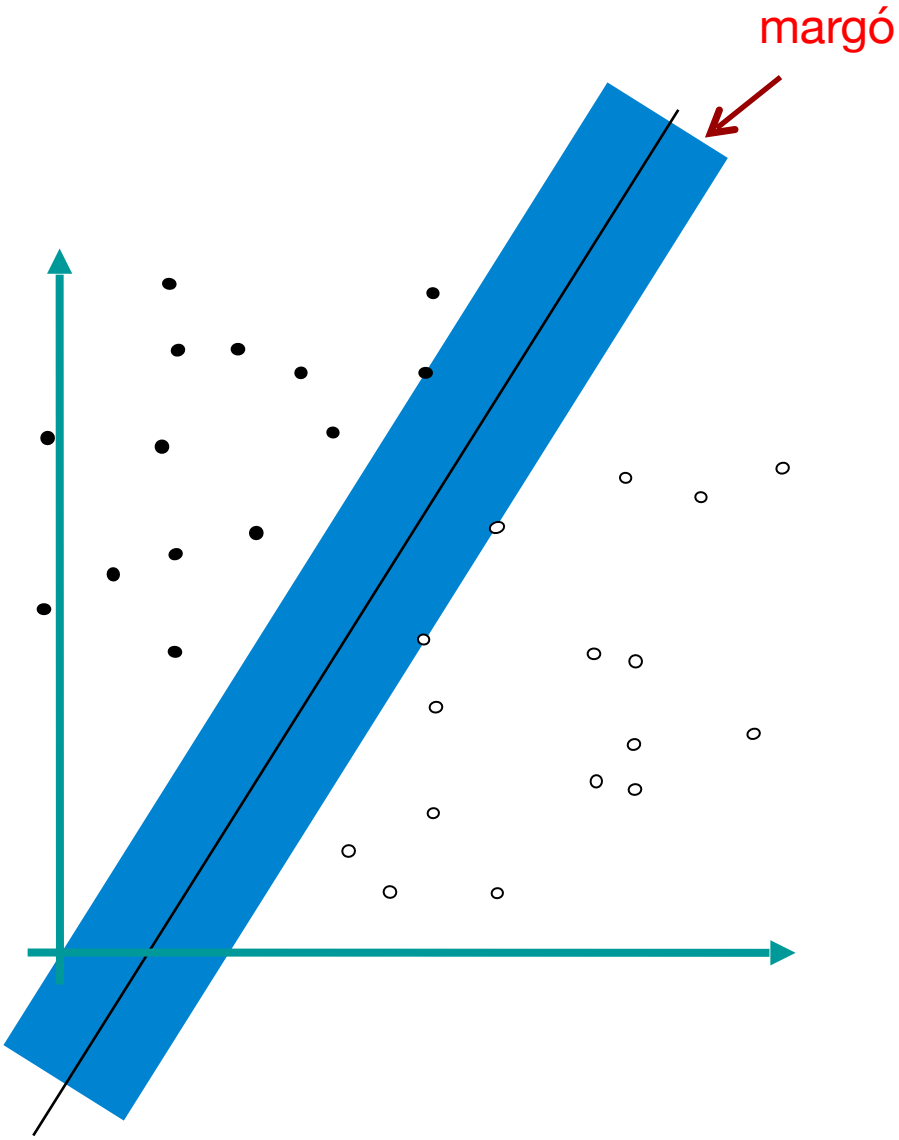
Definíció: Legyen w (X,y) egy lineáris szeparátora. Ebben az esetben a margó a hipersík és bármely minta x_i távolsága

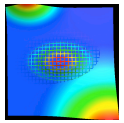
$$min_i \frac{w\, x_i}{|w|}$$

Amennyiben létezik egy ideális megoldás (ahol maximális a margó), a perceptron tanulás konvergál.

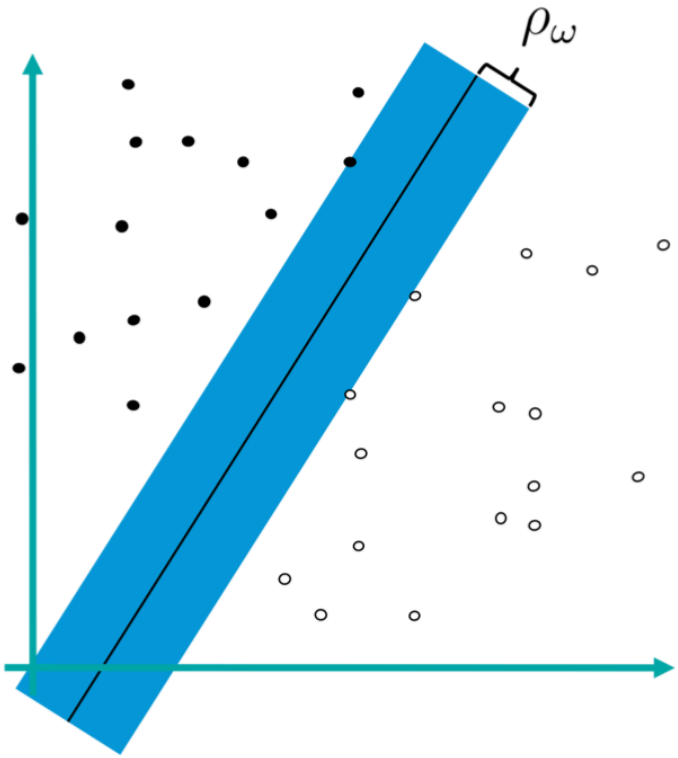


Margó





Maximális margó



Egy modell margója:

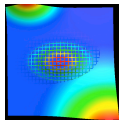
$$\rho_\omega(X) = \min_{x \in X} \frac{|x^T \omega|}{|| \omega ||}$$

Szeretnénk maximalizálni a margót miközben megoldjuk az eredeti problémát, azaz:

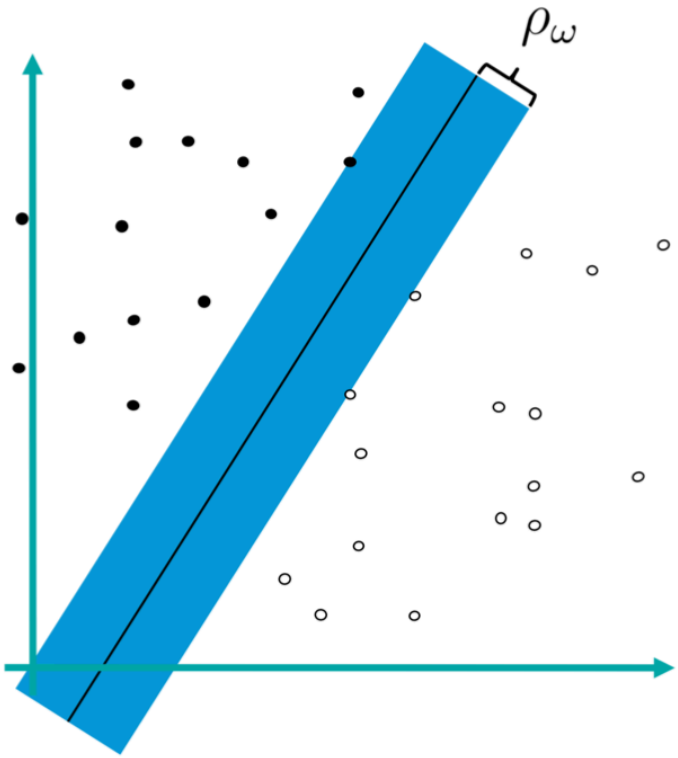
$$\omega* = \arg\max_{\omega \in \Omega} \rho_\omega(X)$$

$$\text{subject to } y_t x_t^T \frac{\omega}{|| \omega ||} > 0, \forall t$$

ahol az osztálycimkék {-1,+1}



Maximális margó

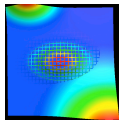


Kihasználva a szigmoid függvény monotonitását, a következő értelmezés is helyes:

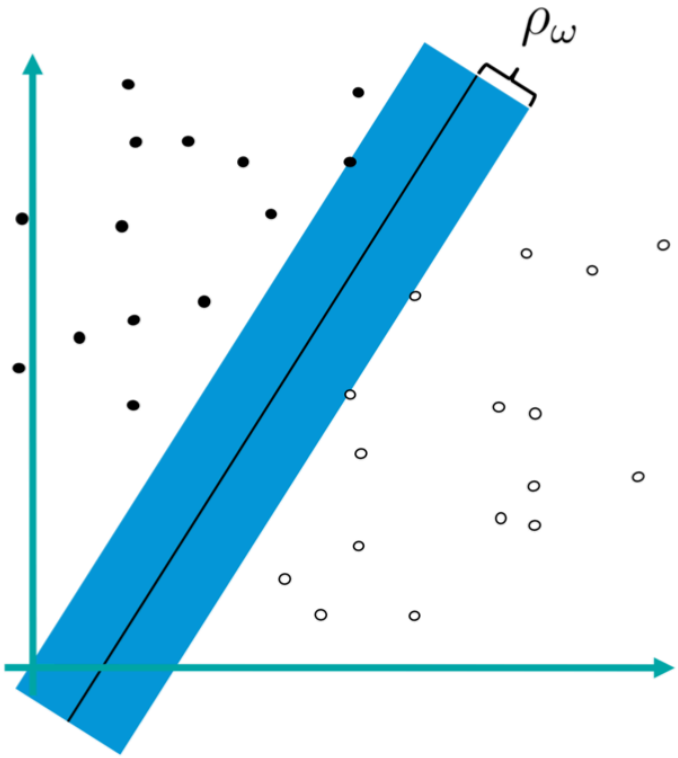
$$\omega* = \arg\max_{\omega\in\Omega} \min_{x\in X} \mid p(x \mid \omega) - 0.5 \mid$$

$$\text{subject to } y_t x_t^T \frac{\omega}{\mid\mid \omega \mid\mid} > 0, \forall t$$

vagyis maximalizáljuk a minimum bizonytalanságot (a valószínűség és az eldöntetlen távolsága)



Maximális margó



Definíció szerint (a hipersíktól vett minimum távolság (megengedünk egyenlőséget!))

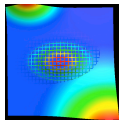
$$y(x^T \frac{\omega}{||\ \omega\ ||}) \geq \rho_\omega$$

a tanulóhalmaz minden (x,y) párjára -> definiálhatunk egy új hipersíkot

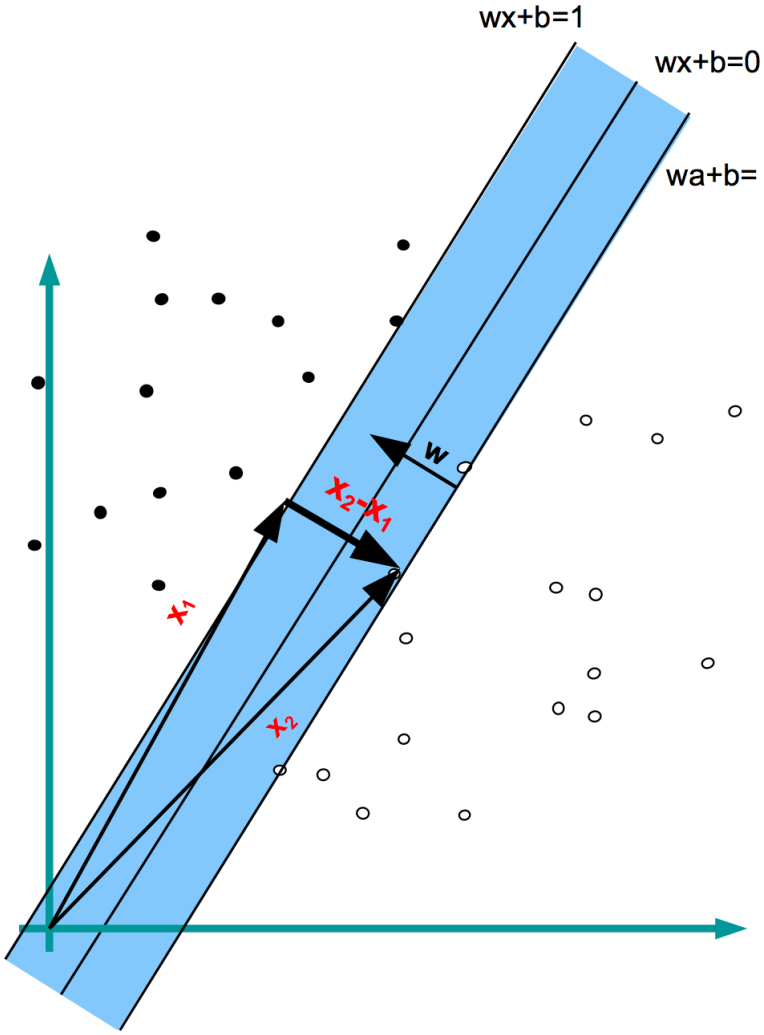
$$\omega' = \frac{\omega}{||\omega|| \rho_\omega}$$

mely minden (x,y)-ra:

$$y(x^T \omega') \geq 1$$



Maximális margó



$$wx_1 + b = 1$$

$$wx_2 + b = -1$$

$$x_2 - x_1 = -n\frac{w}{\|w\|} \longrightarrow x_2 = x_1 - n\frac{w}{\|w\|}$$

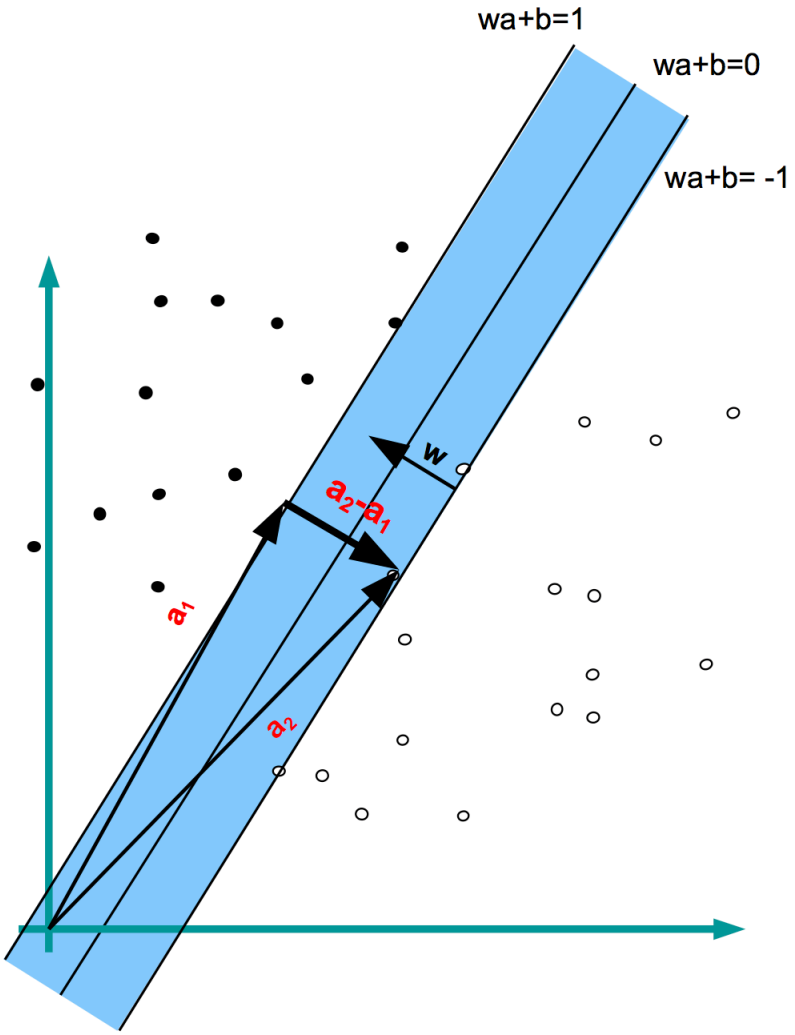
$$wx_2 + b = \left(x_1 - n\frac{w}{\|w\|}\right)w + b = -1$$

$$wx_1 + b = 1$$

$$w\,x_1 + b - n\frac{w}{\|w\|}w = -1$$

$$1 - n\frac{w}{\|w\|}w = -1 \longrightarrow \boxed{n = \frac{2}{\|w\|}}$$

Maximális margó



$$wa_1 + b = 1$$

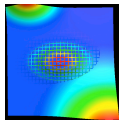
$$wa_2 + b = -1$$

$$a_2 - a_1 = -n\frac{w}{\|w\|} \rightarrow a_2 = a_1 - n\frac{w}{\|w\|}$$

Maximize $n = \frac{2}{\|w\|}$

Minimize $\|w\|$

$$1 - n\frac{w}{\|w\|}w = -1 \rightarrow n = \frac{2}{\|w\|}$$



Maximális margó

Definíció: egy hipersík margója a hipersík és bármely minta minimális távolsága:

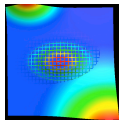
$$\rho_\omega(X) = \min_{x \in X} \frac{|x^T \omega|}{|| \omega ||}$$

Ideális megoldás <-> maximális margó ÉS jó cimkézés:

$$\omega* = \arg\max_{\omega \in \Omega} \rho_\omega(X)$$

$$\text{ahol} \qquad y_t x_t^T \frac{\omega}{|| \omega ||} > 0, \forall t$$

$$y_t \in \{-1, +1\}$$



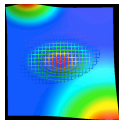
Maximális margó (balról)

A logisztikus regresszióhoz hasonló értelmezés:

$$\omega* = \arg\max_{\omega\in\Omega} \min_{x\in X} \mid p(x \mid \omega) - 0.5 \mid$$

$$\text{ahol} \qquad y_t x_t^T \frac{\omega}{\mid\mid \omega \mid\mid} > 0, \forall t$$

ahol $$p(x \mid \omega) = sigm(x \mid \omega) = \frac{1}{1 + \mathrm{e}^{-(x^T\omega + \omega_0)}}$$



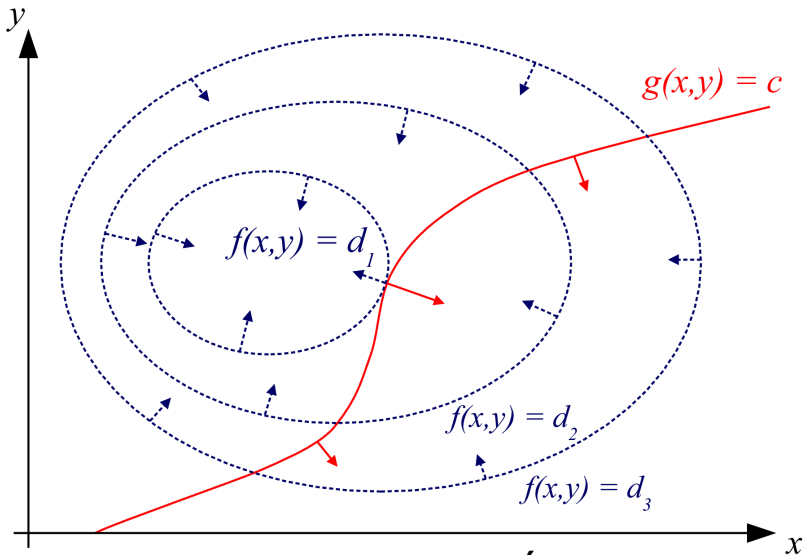
Maximális margó (balról)

A végső (nem soft) optimalizálásunk átformázható

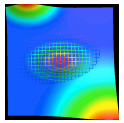
$$\min_\omega \mid\mid \omega \mid\mid^2$$
$$\text{ahol } y_t(x_t^T \omega) \geq 1, \forall t$$

Miért nem lehet közvetlenü megoldani?

Megoldás: Lagrange duális!

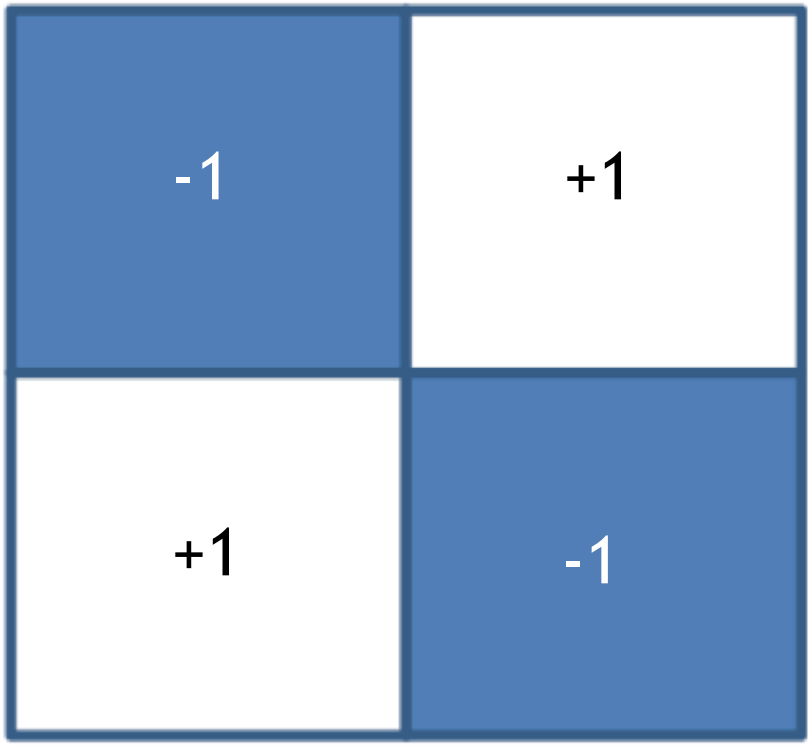


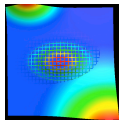
Ábra: wiki



Nem lineárisan szeparálható problémák

Lehetséges-e lineáris szeparálót találni a lenti adathalmazhoz?





Polinomiális szeparátor

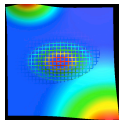
Feltételezésünk szerint létezik egy p(x) D-ed fokú polinom, mely p(x)>0 akkor és csak akkor ha x pozitív cimkéjű.

Ha D-ed fokú a polinomunk akkor minden $(i_1,i_2,\ldots,i_d)$-re igaz , hogy

$$i_1 + i_2 + \cdots + i_d \leq D$$

Azaz a monomok formája

$$x_1^{i_1} x_2^{i_2} \cdots x_d^{i_d}$$



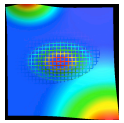
Polinomiális szeparáló

Ha a monomok együtthatói $w_{i1,i2,\ldots,id}$, a polinomunk a következő alakot veszi fel

$$p(x_1, x_2, \ldots, x_d) = \sum_{\substack{i_1,i_2,\ldots,i_d \\ i_1+i_2+\cdots+i_d \leq D}} w_{i_1,i_2,\ldots,i_d} x_1^{i_1} x_2^{i_2} \cdots x_d^{i_d}$$

Már kis D mellett is lehet nagy az együtthatók száma!

$$\sum_{k=1}^{D} \binom{d+k-1}{k} + 1$$



Polinomiális szeparáló

Példa: legyen d és D is 2.
Ebben az esetben a monomok száma 6,

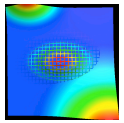
I_1, i_2 a lehetséges értékei $\{(0,0),(1,0),(0,1),(2,0),(1,1),(0,2)\}$

Legyen (0,0) a “bias” (b), a polinomunk végső formája

$$p(x_1, x_2) = b + w_{10}x_1 + w_{01}x_2 + w_{11}x_1x_2 + w_{20}x_1^2 + w_{02}x_2^2$$

Minden $\mathbf{x}_i$ tanulópontra

$b + w_{10}x_{i1} + w_{01}x_{i2} + w_{11}x_{i1}x_{i2} + w_{20}x_{i1}^2 + w_{02}x_{i2}^2 > 0$ ha a címke $y_i = +1$



Polinomiális szeparáló

A fenti eset egy véges transzformációnak tekinthető.

Legyen minden $i_1,i_2,\ldots,i_d$ összege legfeljebb D, akkor a transzformációnk

$$\varphi(\mathbf{x})$$ R^d -> R^m

$$\mathbf{a_i} = (x_1, x_2, \ldots, x_d) \quad \text{->} \quad x_1^{i_1} x_2^{i_2} \cdots x_d^{i_d}$$

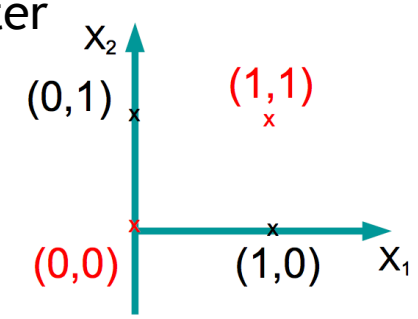
ha d=2 és D=2: $$\varphi(\mathbf{x}) = (x_1, x_2, x_1^2, x_1 x_2, x_2^2)$$

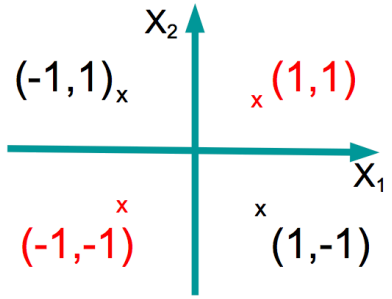
ha d=3 és D=2:

$$\varphi(\mathbf{x}) = (x_1, x_2, x_3, x_1 x_2, x_1 x_3, x_2 x_3, x_1^2, x_2^2, x_3^2)$$

Polinomiális szeparáló

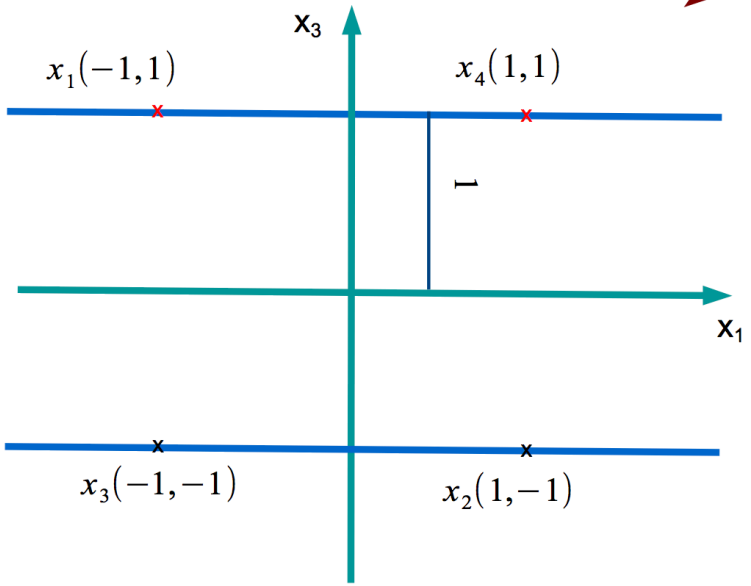
Eredeti tér





$\Phi(X)$

Transzformált tér

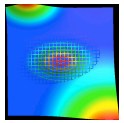


$x_{11}=-1\,,\,x_{13}=1$ $\quad y_1=-1$

$x_{21}=1\,,\,x_{23}=-1$ $\quad y_2=+1$

$x_{31}=-1\,,\,x_{33}=-1$ $\quad y_3=+1$

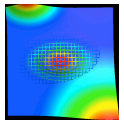
$x_{41}=1\,,\,x_{43}=1$ $\quad y_4=-1$



Polinomiális kernel

Illesszük vissza az eredeti problémánkba a polinomiális transzformációnkat.

Álljunk meg egy pillanatra! El tudjuk kerülni a transzformált vektorok kiszámolását?



Support vector machine

Lemma: konvex problémánk megoldása felírható a minták lineáris kombináltjaként

Ebből következik hogy a norma

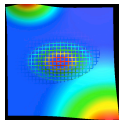
$$||\omega||^2 = ((\sum \alpha_i x_i)^T (\sum \alpha_i x_i)) = \sum \alpha_i \alpha_j x_i x_j$$

Illetve minden tanulópontra a következő igaz

$$y_i(\omega^T x) = y_i \sum \alpha_i x_i x = y_i \sum \alpha_i k(x_i, x)$$

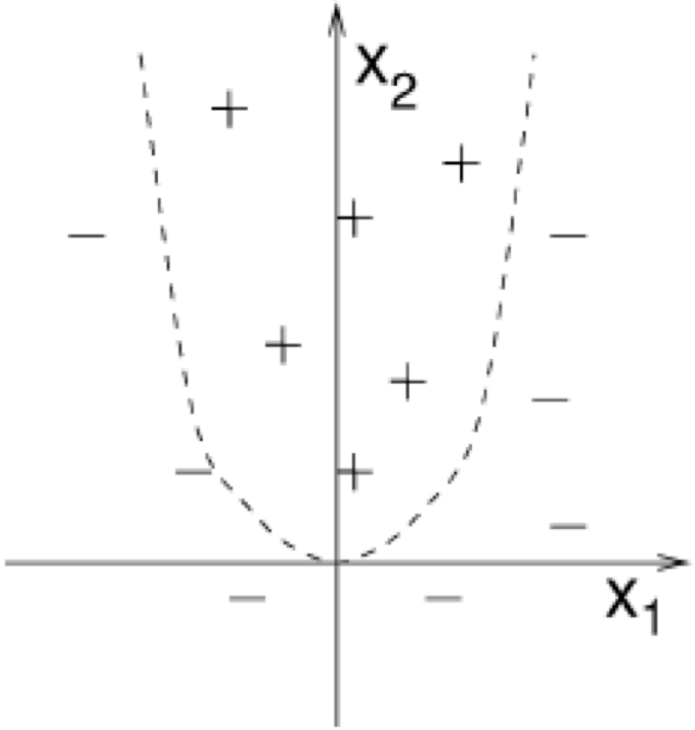
ahol a k(x,x) függvényt kernel (mag) függvénynek nevezzük. A nem 0-a együtthatójú tanulóelemek a support vektorok (innen a név).

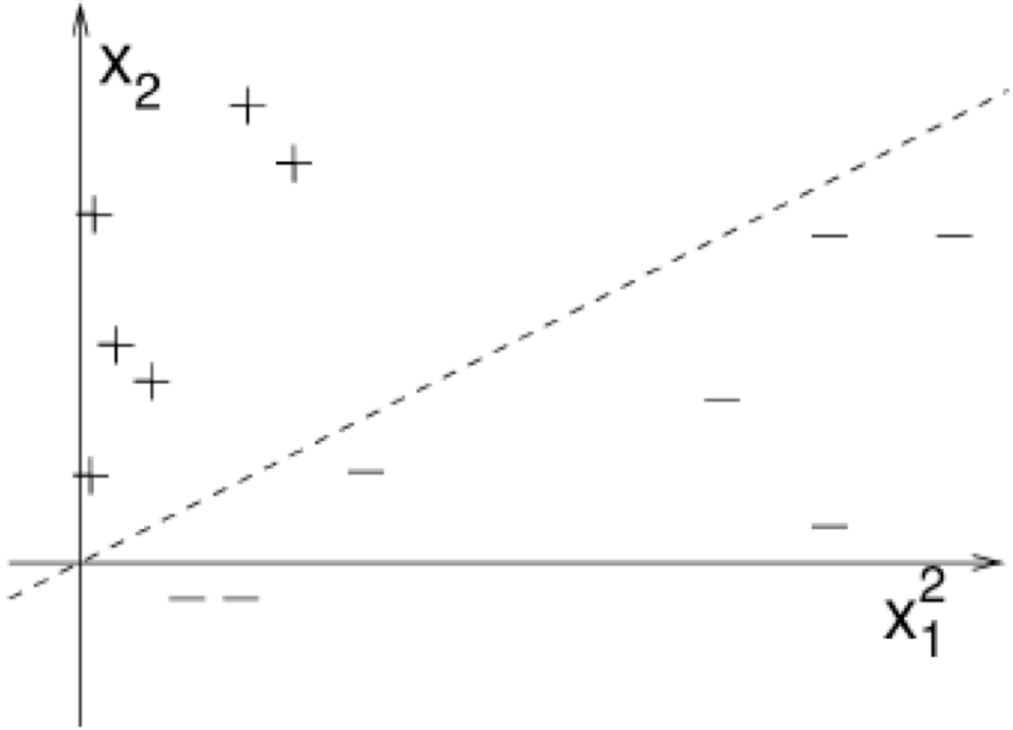
Órai feladat 1: Készítsünk polinomiális szeparálót a belső szorzatok ismeretével!

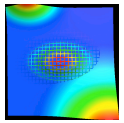


Polinomiális kernel

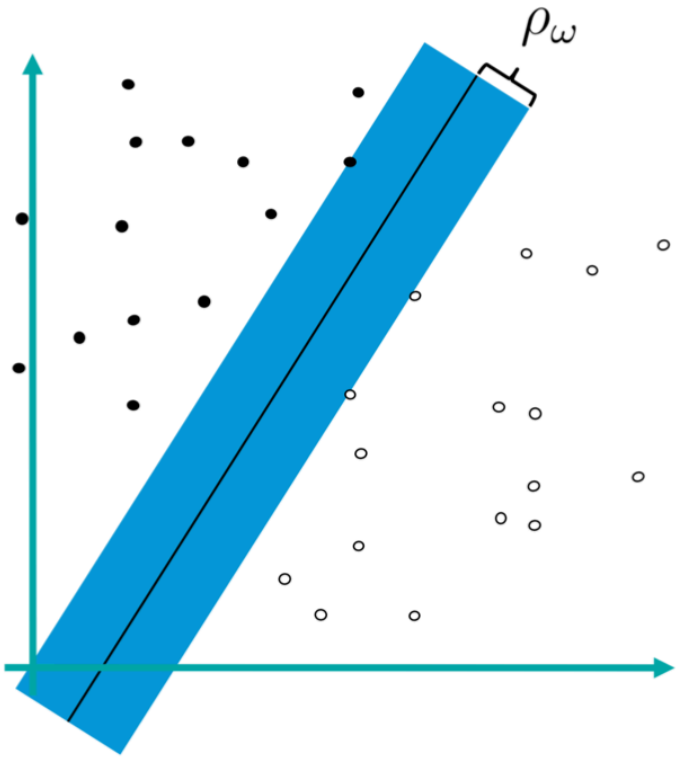
$$\Phi(x)=(x_1^2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{(2)}x_1x_2, 1)$$







Maximális margó

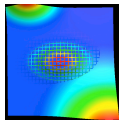


Mindezek alapján az eredeti maximalizálási probléma megegyezik a következő minimalizálási problémával:

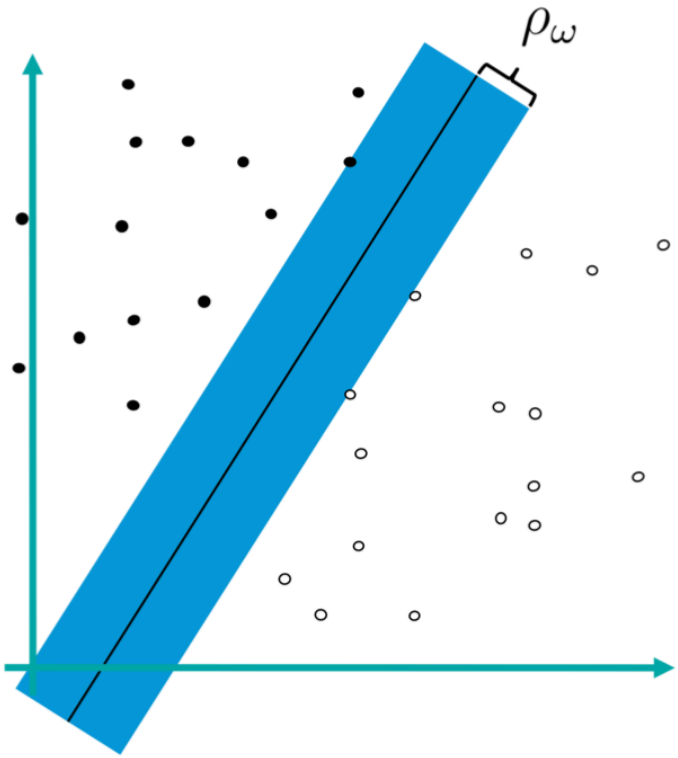
$$\text{minimize}_\omega \frac{1}{2} \mid\mid \omega \mid\mid^2$$

$$\text{subject to } y_t(x_t^T \omega) \geq 1, \forall t$$

Miért ½ ?
Miért négyzet?



Maximális margó

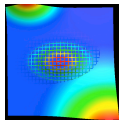


Ugyan a problémánk konvex, kvadratikus, mégse tudjuk megoldani közvetlenül... a feltételek miatt

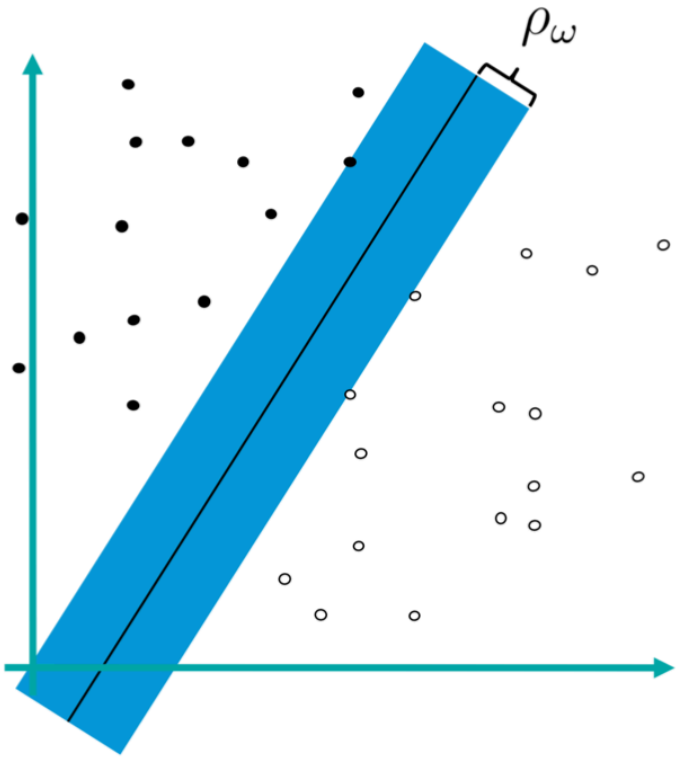
Szerencsére a feltételek is folytonosan differenciálható függvények -> Lagrange!

Formálisan, legyenek $\alpha_t \geq 0, \forall t$ a Lagrange probléma primális változói (Lagrangian multipliers) és a Lagrange függvény:

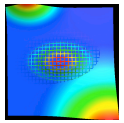
$$L(\omega, \alpha) = \frac{1}{2} \mid\mid \omega \mid\mid^2 - \sum_{t=1}^{T} \alpha_t (y_t (x_t^T \omega) - 1)$$



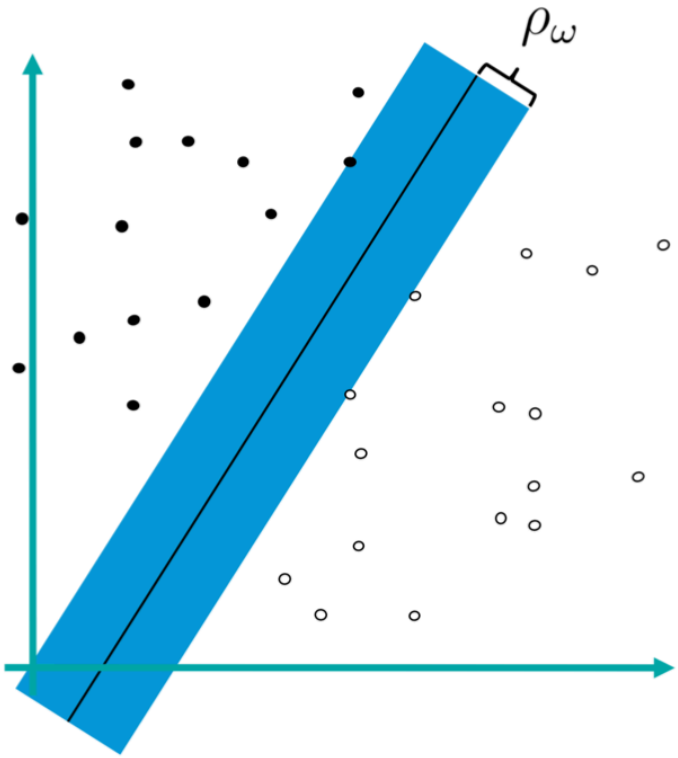
Maximális margó



De miért Lagrange?



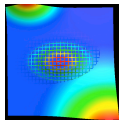
Maximális margó



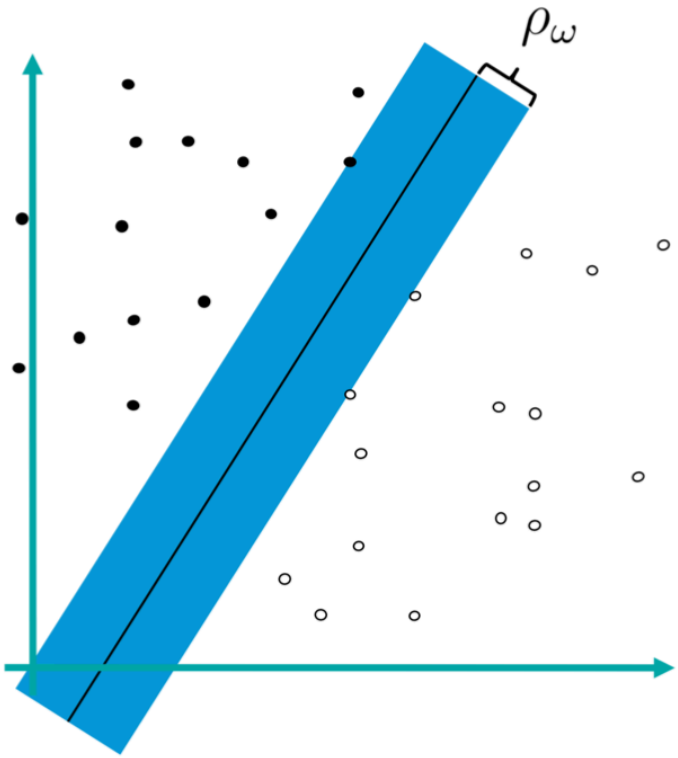
De miért Lagrange?

A Lagrange függvény parciális deriváltja a normál vektorra nézve pontosan ott lesz nulla ahol az eredeti optimalizációnk optimuma található (majd minden esetben...)

Akkor számoljuk ki!



Maximális margó



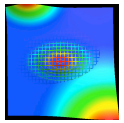
De miért Lagrange?

A Lagrange függvény parciális deriváltja a normál vektorra nézve pontosan ott lesz nulla ahol az eredeti optimalizációnk optimuma található (majd minden esetben...)

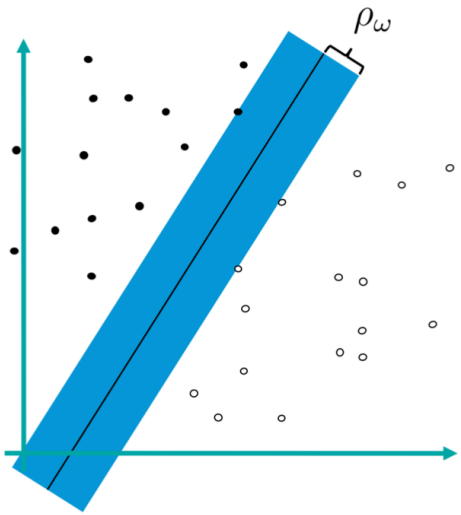
Akkor számoljuk ki!

$$\frac{\partial L(\omega, \alpha)}{\partial \omega_i} = \omega_i - \sum_{t=1}^{T} y_t \alpha_t x_{ti} = 0$$

Stacionárius pont: ahol a derivált nulla.



Maximális margó



Ismétlés:

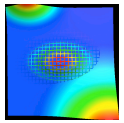
A normál vektor felírható a tanulópontok lineáris kombinációjaként (miért is?)

$$\omega = \sum_{t=1}^{T} \alpha_t y_t x_t$$

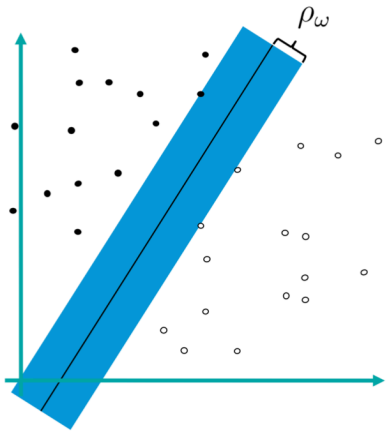
S most tegyük mindet a tálba:

$$L(\omega, \alpha) = L(\alpha) = \frac{1}{2} \sum_{i=1}^{T} \sum_{j=1}^{T} \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^{T} \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{t=1}^{T} \alpha_t$$

$$= \sum_{t=1}^{T} \alpha_t - \frac{1}{2} \sum_{i=1}^{T} \sum_{j=1}^{T} \alpha_i \alpha_j y_i y_j x_i^T x_j$$



Maximális margó



A mostmár véglegesnek tekinthető optimalizációnk...

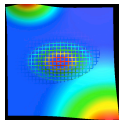
Megjegyzés, a második feltétel a bias ...

Ez azért nem tűnik egyszerűnek... Hogyan lehet megoldani?

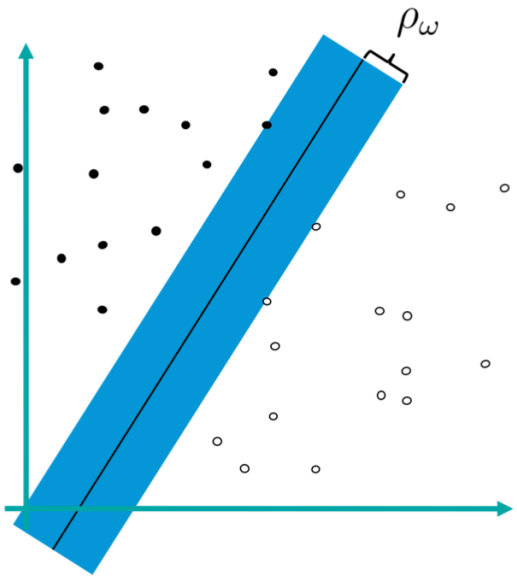
$$\text{maximize}_\alpha L(\alpha) = \sum_{t=1}^{T} \alpha_t - \frac{1}{2} \sum_{i=1}^{T} \sum_{j=1}^{T} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$subject\ to\ \alpha_i \geq 0, \forall i$$

$$\sum_{t=1}^{T} \alpha_t y_t = 0, \forall t$$



Maximális margó



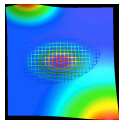
Karush-Kuhn-Tucker feltételek, függetlenül feltaláltak:

Karush 1939 - Master tézis
Kuhn-Tucker 1951 – függetlenül

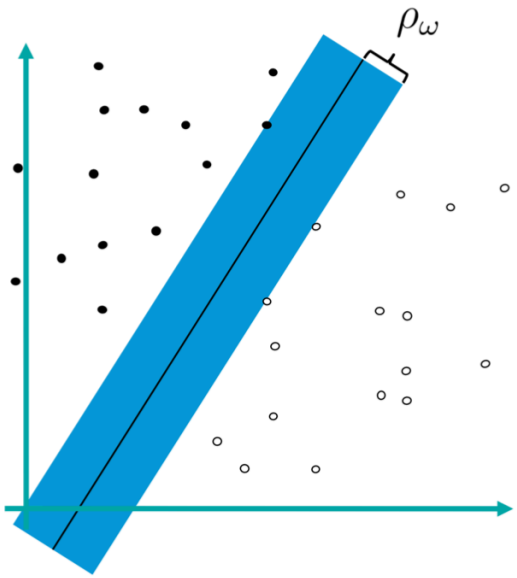
Az optimális megoldás pozitív Lagrange szorzókat tartalmaz és

$$\alpha_i(y_i(x_i^T\omega) - 1) = 0, \forall i$$

Hogyan lehetne ezt értelmezni?



Maximális margó



Karush-Kuhn-Tucker feltételek, függetlenül feltaláltak:

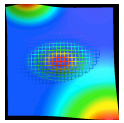
Karush 1939 - Master tézis
Kuhn-Tucker 1951 – függetlenül

Az optimális megoldás pozitív Lagrange szorzókat tartalmaz és

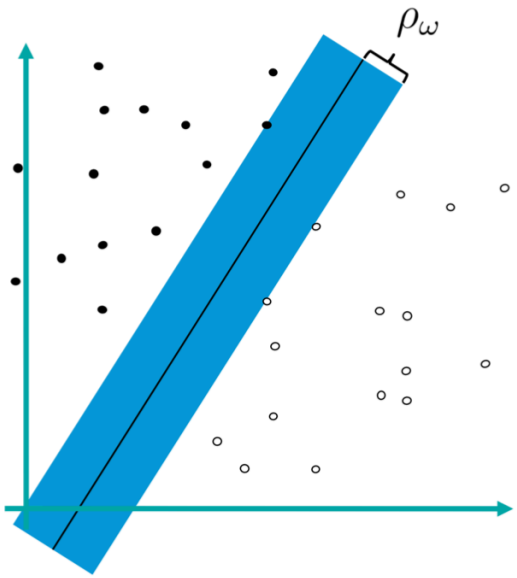
$$\alpha_i(y_i(x_i^T\omega) - 1) = 0, \forall i$$

Hogyan lehetne ezt értelmezni?

Cortes-Vapnik (1995) nem nulla szorzóval (KKT: pozitív) rendelkező tanulópontok a "Support Vector"-ok SVM :)



Maximális margó



Az optimális megoldás pozitív Lagrange szorzókat tartalmaz és

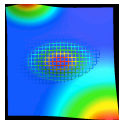
$$\alpha_i(y_i(x_i^T\omega) - 1) = 0, \forall i$$

Hogyan lehetne ezt értelmezni?

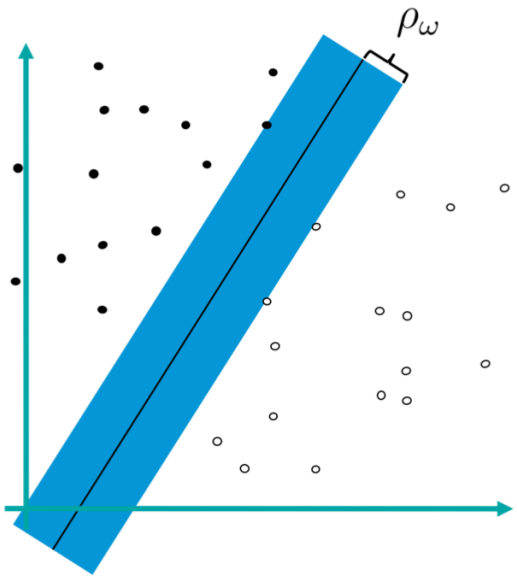
Cortes-Vapnik (1995) nem nulla szorzóval (KKT: pozitív) rendelkező tanulópontok a "Support Vector"-ok SVM :)

Következmény ("haszontalan" pontok):

$$\omega = \sum_{t=1}^{T} \alpha_t x_t = \sum_{x_i \in SV} \alpha_i x_i$$



Maximális margó



"Soft-margin": valamilyen szinten megengedünk pontokat a margón belül... "1-Norm soft margin":

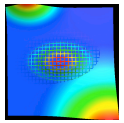
$$\text{minimize } \frac{1}{2} \mid\mid \omega \mid\mid^2 + C \sum_{t=1}^{T} \xi_i$$

$$\text{subject to } y_i(x_i^T \omega) \geq 1 - \xi_i, \forall i$$

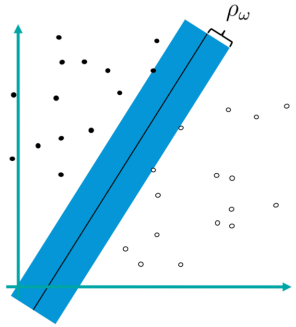
$$\xi_i \geq 0, \forall i$$

ahol C egy előre meghatározott konstans (ha nulla -> "hard margin")

Újra Lagrange függvény :)



Maximális margó

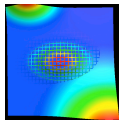


Lagrange függvény:

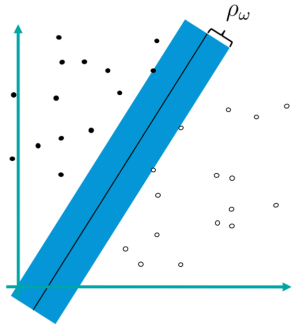
$$L(\omega, \alpha, \beta) = \frac{1}{2} \mid\mid \omega \mid\mid^2 + C \sum_{t=1}^{T} \xi_i - \sum_{t=1}^{T} \alpha_t (y_t (x_t^T \omega) - 1 + \xi_t) - \sum_{t=1}^{T} \beta_t \xi_t$$

Ahol beta egy újabb Lagrange szorzó.

Mi a következő lépés?



Maximális margó



Lagrange függvény:

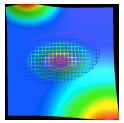
$$L(\omega, \alpha, \beta) = \frac{1}{2} \mid\mid \omega \mid\mid^2 + C \sum_{t=1}^{T} \xi_i - \sum_{t=1}^{T} \alpha_t (y_t (x_t^T \omega) - 1 + \xi_t) - \sum_{t=1}^{T} \beta_t \xi_t$$

Ahol beta egy újabb Lagrange szorzó.

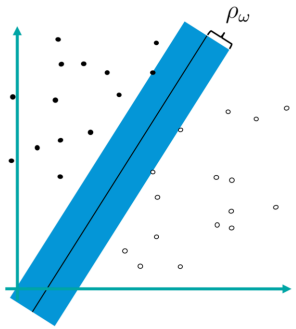
Mi a következő lépés? A derivált!

$$\frac{\partial L(\omega, \xi, \alpha, \beta)}{\partial \omega_i} = \omega_i - \sum_{t=1}^{T} \alpha_t y_t x_{ti} = 0$$

$$\frac{\partial L(\omega, \xi, \alpha, \beta)}{\partial \xi_i} = C - \alpha_i - \beta_i = 0, \forall i$$



Maximális margó



Deriváltak:

$$\frac{\partial L(\omega,\xi,\alpha,\beta)}{\partial \omega_i} = \omega_i - \sum_{t=1}^{T} \alpha_t y_t x_{ti} = 0$$

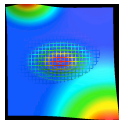
$$\frac{\partial L(\omega,\xi,\alpha,\beta)}{\partial \xi_i} = C - \alpha_i - \beta_i = 0, \forall i$$

A normál vektor szerinti gradiens ugyanaz mint volt nem "soft" esetben! (Cortes-Vapnik 1995)

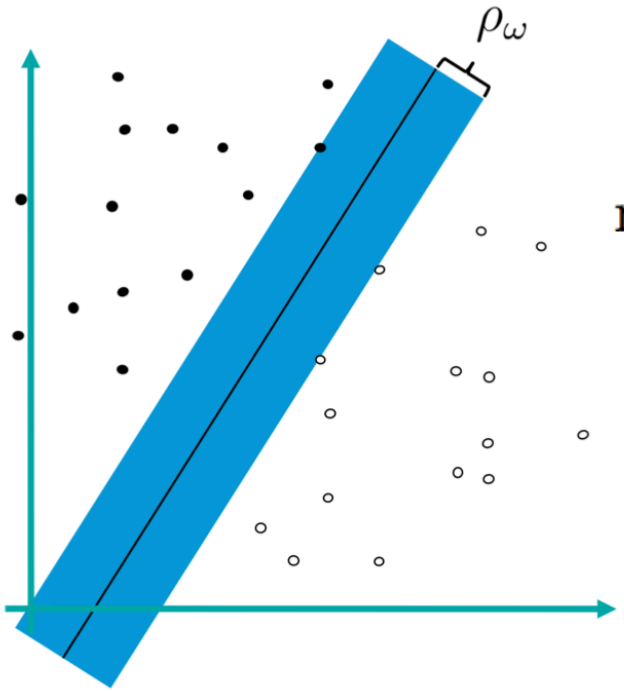
S mivel tudjuk a KKT feltételek miatt, hogy mindkét szorzó pozitív:

$$\alpha_i(y_i x_i^T \omega - 1 + \xi_i) = 0, \forall i$$

$$\xi_i(C - \alpha_i) = 0, \forall i.$$



Maximális margó



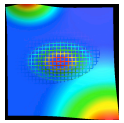
Ebből következik hogy C felülről korlátozza alfát, illetve alfa pozitív:

$$\text{maximize } W(\alpha) = \sum_{t=1}^{T} \alpha_t - \frac{1}{2} \sum_{i=1}^{T} \sum_{j=1}^{T} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$\text{subject to } 0 \leq \alpha_i \geq C, \forall i.$$

így a derivált:

$$\frac{\partial W(\alpha)}{\partial \alpha_i} = 1 - y_i \sum_{t=1}^{T} \alpha_j y_j x_i^T x_j$$



Maximal margin (a bit of recap)

Amennyiben a belső szorzatot lecseréljük bármilyen transzformált térbeli belső szorzatra:

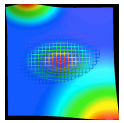
Algorithm 1-Norm Soft Margin SVM

Given a training set $X = \{x_1, .., x_T\}$ with $x_i \in \mathbb{R}^d, \forall i$, a positive real valued constant C, a positive real valued learning rate η and a kernel function $K(x, y) = \phi(x)^T \phi(y)$

$\alpha \leftarrow 0$
repeat
for $i = 1$ to T

1. $\alpha_i^{new} \leftarrow \alpha_i^{old} + \eta \frac{\partial W(\alpha)}{\partial \alpha_i} = \alpha^{old} + \eta(1 - y_i \sum_{t=1}^{T} \alpha_t y_t K(x_t, x_i))$
2. if $\alpha_i < 0$ then $\alpha_i \leftarrow 0$
   else
   if $\alpha_i > C$ then $\alpha_i \leftarrow C$

end for
until we reach a stopping criterion
return α



Vapnik-Chervonenkis tételek

Lady (Dr. Muriel Bristol-Roach) :“by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup"

R. A. Fisher és 8 csésze -> 4/4

Fisher megkérte "a Lady"-t hogy válassza ki azt a négy csészét, melybe a tejet öntötte először ízlelés alapján

Mennyi a valószínűsége, hogy 0,1,2,3,4 helyes csészét választott?

Vapnik-Chervonenkis tételek

Lady (Dr. Muriel Bristol-Roach) :“by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup"

R. A. Fisher és 8 csésze -> 4/4

Fisher megkérte "a Lady"-t hogy válassza ki azt a négy csészét, melybe a tejet öntötte először ízlelés alapján

Mennyi a valószínűsége, hogy 0,1,2,3,4 helyes csészét választott?

Összesen 70 eset.

Nulla helyes: 1
Egy helyes: 16
Két helyes: 36
Három helyes: 16
Négy helyes: 1

Brief intro to VC theorem

Lady (Dr. Muriel Bristol-Roach) :“by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup"

R. A. Fisher prepared 8 cups -> 4/4

He asked the Lady to choose 4 cups in which she thinks the milk was added first.

What is the probability of having 0,1,2,3,4 correct answers?

Overall: 70 cases

Out of them:

Having zero correct answer: 1
Having one correct answer: 16
Having two correct answers: 36
Having three correct answers: 16
Having four correct answers: 1