Üzleti Elemzések Módszertana

6. Előadás: Tartó vektor gépek

Kuknyó Dániel
Budapesti Gazdasági Egyetem

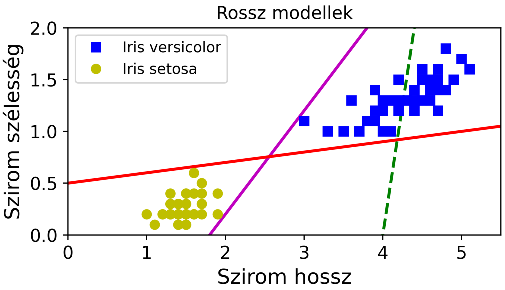
2023/24
2.félév

A vektorgépek mögötti intuíció

Az SVM modellek minden esetben a **legszélesebb utat** keresik egy adathalmaz két osztálya között. Az út közepén húzódik a **döntési határ** és **az út két oldalán fekszenek a margók**.

Az úton kívül hozzáadott mintaegyedek nem befolyásolják a döntési határt, tehát az utat a csoportok szélén lévő mintaegyedek tartják.

A margó egyik oldalán A, a másikon pedig B osztály lesz a predikció.

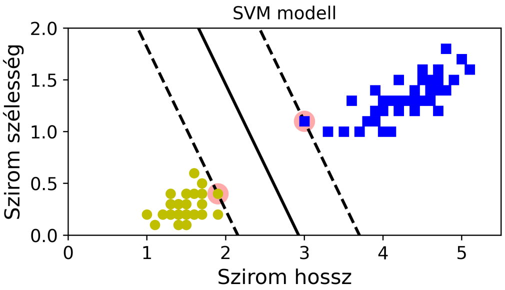


A vektorgépek mögötti intuíció

Az SVM modellek minden esetben a **legszélesebb utat** keresik egy adathalmaz két osztálya között. Az út közepén húzódik a **döntési határ** és **az út két oldalán fekszenek a margók**.

Az úton kívül hozzáadott mintaegyedek nem befolyásolják a döntési határt, tehát az utat a csoportok szélén lévő mintaegyedek tartják.

A margó egyik oldalán A, a másikon pedig B osztály lesz a predikció.



Adat előkészítés

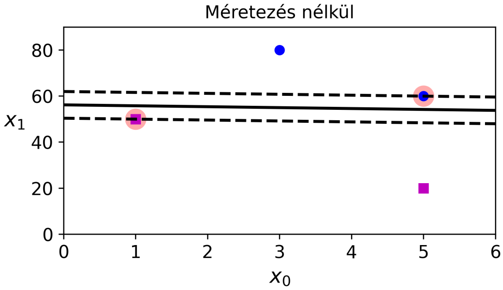
A tartó vektor gépek nagyon szenzitívek az adatok méretezésére.

Normalizálás

A normalizálás (vagy méretezés) célja, hogy az adatkészlet különböző változóit egységes mértékrendszerbe hozza.

Ennek következménye, hogy a predikcióhoz minden változó egyenlő súllyal hozzájáruljon.

Érdemes megfigyelni az út változását méretezés előtt és után.



Adat előkészítés

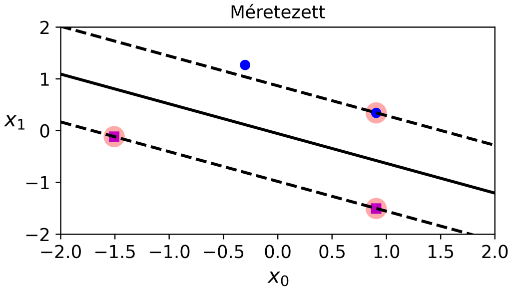
A tartó vektor gépek nagyon szenzitívek az adatok méretezésére.

Normalizálás

A normalizálás (vagy méretezés) célja, hogy az adatkészlet különböző változóit egységes mértékrendszerbe hozza.

Ennek következménye, hogy a predikcióhoz minden változó egyenlő súllyal hozzájáruljon.

Érdemes megfigyelni az út változását méretezés előtt és után.



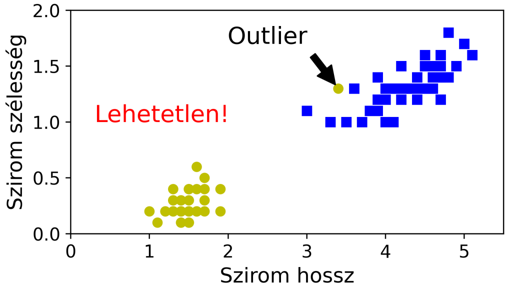
Keménymargós osztályozás

Keménymargós osztályozás

Amennyiben áll az a feltétel, hogy minden mintaegyednek az úton kívül kell esnie, az osztályozás **keménymargós**.

Ez csak akkor lehetséges, amikor az adatok lineárisan szeparálhatóak. Ezenkívül a kiugró adatpontok képesek irracionálisan torzítani a modell határait.

Ennek kiküszöbölésére olyan modellre van szükség, amelyik engedélyezi a **margósértéseket**.



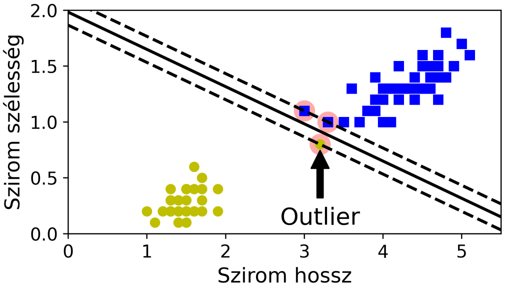
Keménymargós osztályozás

Keménymargós osztályozás

Amennyiben áll az a feltétel, hogy minden mintaegyednek az úton kívül kell esnie, az osztályozás **keménymargós**.

Ez csak akkor lehetséges, amikor az adatok lineárisan szeparálhatóak. Ezenkívül a kiugró adatpontok képesek irracionálisan torzítani a modell határait.

Ennek kiküszöbölésére olyan modellre van szükség, amelyik engedélyezi a **margósértéseket**.



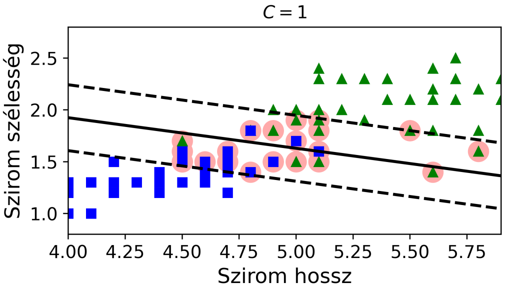
Lágymargós osztályozás

A margósértések és a rosszul generalizáló modellek közötti optimum keresésére hivatottak a lágymargós SVM modellek.

Lágymargós osztályozás

Olyan SVM modell, amely **engedélyezi a margósértéseket. A margó keménységét a C hiperparaméter szabályozza.**

A kisebb C érték szélesebb úthoz, de több margósértéshez vezet. A nagyobb C érték pedig pedig kevesebb margósértést, de rosszabbul generalizáló modellt eredményez.



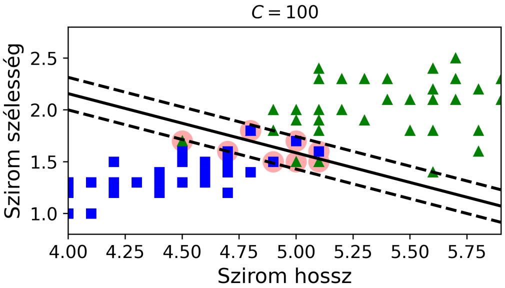
Lágymargós osztályozás

A margósértések és a rosszul generalizáló modellek közötti optimum keresésére hivatottak a lágymargós SVM modellek.

Lágymargós osztályozás

Olyan SVM modell, amely **engedélyezi a margósértéseket. A margó keménységét a C hiperparaméter szabályozza.**

A kisebb C érték szélesebb úthoz, de több margósértéshez vezet. A nagyobb C érték pedig pedig kevesebb margósértést, de rosszabbul generalizáló modellt eredményez.

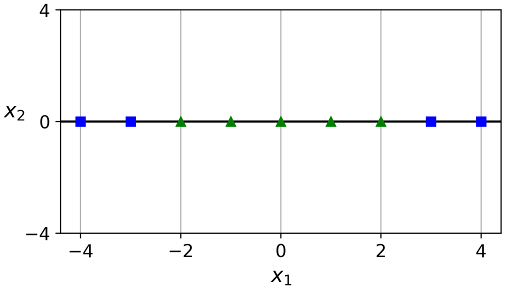


Nemlineáris SVM osztályozás

Habár az SVM modellek jól teljesítenek lineárisan szeparálható adatok esetén, a valóságban ezek az adathalmazok nagyon ritkának számítanak.

Egy módja a nemlineáris SVM osztályozásnak, ha a meglévő változókra **egy magasabb dimenziójú térben történik az osztályozás**.

Ebben az esetben a magasabb dimenziós tér a **négyzetes transzformációja** az x_1 változónak.

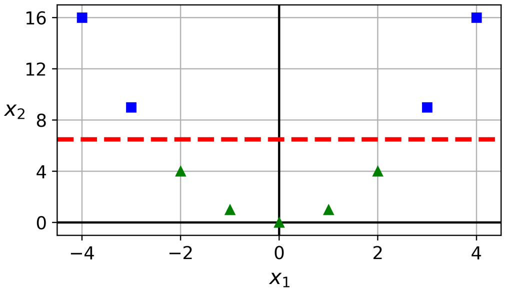


Nemlineáris SVM osztályozás

Habár az SVM modellek jól teljesítenek lineárisan szeparálható adatok esetén, a valóságban ezek az adathalmazok nagyon ritkának számítanak.

Egy módja a nemlineáris SVM osztályozásnak, ha a meglévő változókra **egy magasabb dimenziójú térben történik az osztályozás**.

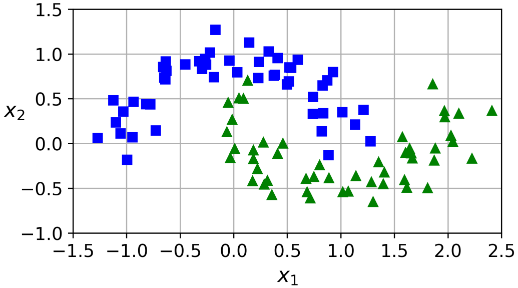
Ebben az esetben a magasabb dimenziós tér a **négyzetes transzformációja** az x_1 változónak.



Nemlineáris SVM a `make-moons` adathalmazon

A transzformációs csővezeték a következő:

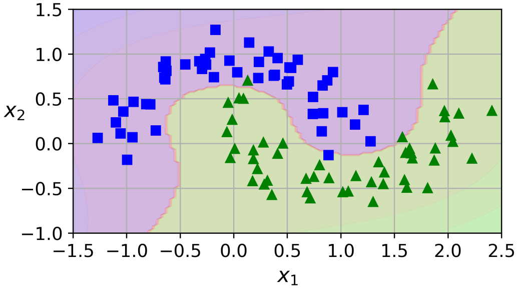
1. Polinomikus jellemzők felvétele
2. Jellemzők normalizálása
3. SVM osztályozó futtatása



Nemlineáris SVM a `make-moons` adathalmazon

A transzformációs csővezeték a következő:

1. Polinomikus jellemzők felvétele
2. Jellemzők normalizálása
3. SVM osztályozó futtatása



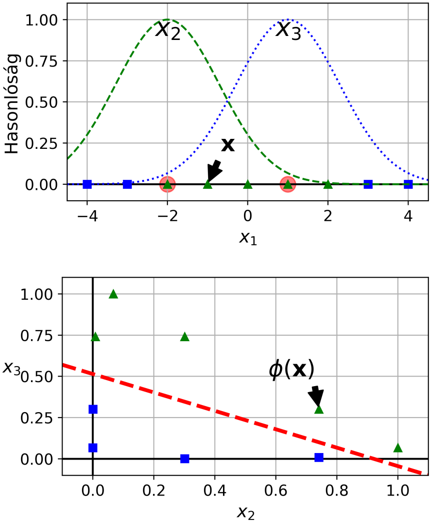
Hasonlósági függvények kernelként

Hasonlósági függvény

Azt reprezentálja, hogy egy adott lekérdezési pont mennyire hasonlít egy előre meghatározott **tájékozódási ponthoz**.

A példában az előző, 1D adathalmazhoz választott két tájékozódási pont $x_1 = -2$ és $x_2 = 1$. A hasonlósági függvény pedig a Gauss-i radiális bázis függvény $\gamma = 0.3$ paraméterrel:

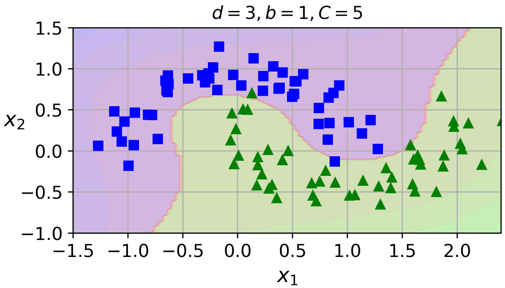
$$\phi_\gamma(x, \ell) = exp\left(-\gamma \left|x - \ell\right|^2\right)$$



Polinomikus függvények kernelként

Polinomikus jellemzők hozzáadása alacsony polinomikus szinten nem képes komplex adathalmazokkal dolgozni.

Magas szinten viszont nagyon lassúvá teheti a modellt, mivel minden polinomikus szinten egy külön transzformációt jelent a jellemzőkre.



Polinomikus függvények kernelként

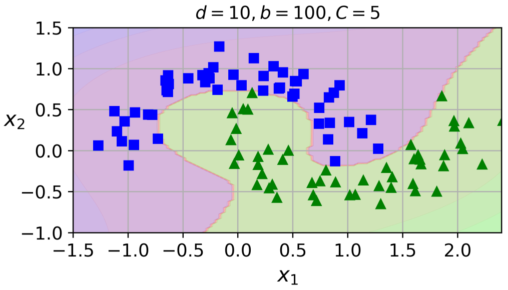
Polinomikus kernel

A polinomikus kernel transzformáció lineárisan nem szeparálható adathalmazokat transzformál olyan magasabb dimenziós térbe, ahol már lineárisan szeparálhatóvá válnak:

$$K(x, y) = ((x \cdot y) + b)^d$$

Ahol:

- b: A konstans torzítás
- d: A transzformáció polinomikus foka



A kernel trükk

Egy φ másodrendű polinomikus leképezés 2D tanító halmazra 3D eredményt ad:

$$\varphi(x) = \varphi\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{pmatrix}$$

Ezt a másodrendű leképezést alkalmazva, majd a vektorok belső szorzatait kiszámolva:

$$\begin{pmatrix} a_1^2 \\ \sqrt{2}a_1a_2 \\ a_2^2 \end{pmatrix}^T \begin{pmatrix} b_1^2 \\ \sqrt{2}b_1b_2 \\ b_2^2 \end{pmatrix} = a_1^2b_1^2 + 2a_1b_1a_2b_2 + a_2^2b_2^2 = \left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}^T \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}\right)^2 = \left(a^Tb\right)^2$$

Látható, hogy a vektorok belső szorzata megegyezik az eredeti vektorok belső szorzatának négyzetével:

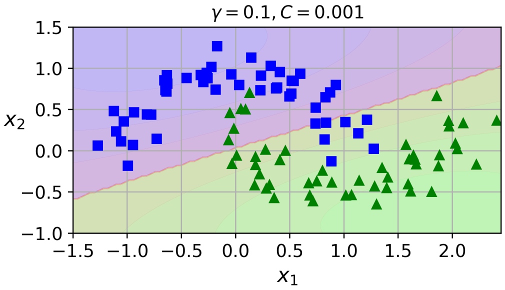
$$\varphi(a)^T\varphi(b) = \left(a^Tb\right)^2$$

Radiális bázis függvények kernelként

Ahogy a polinomikus jellemzők esetében, úgy **a hasonlósági jellemzők esetén is működik a kernel trükk**.

A diagramon különböző γ és C értékekkel tanított modellek láthatóak.

A γ növelése szűkebb haranggörbékkel végzi az illesztést, ezért minden egyed hasonlósági tartománya kisebb. A γ csökkentése nagyobb befolyási teret ad a pontoknak.

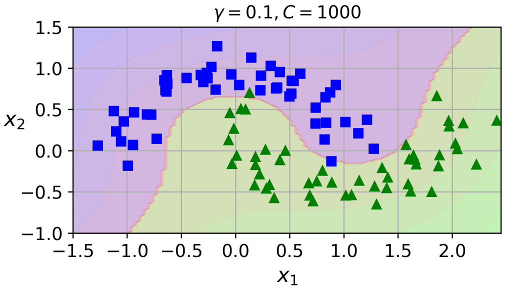


Radiális bázis függvények kernelként

Ahogy a polinomikus jellemzők esetében, úgy **a hasonlósági jellemzők esetén is működik a kernel trükk**.

A diagramon különböző γ és C értékekkel tanított modellek láthatóak.

A γ növelése szűkebb haranggörbékkel végzi az illesztést, ezért minden egyed hasonlósági tartománya kisebb. A γ csökkentése nagyobb befolyási teret ad a pontoknak.

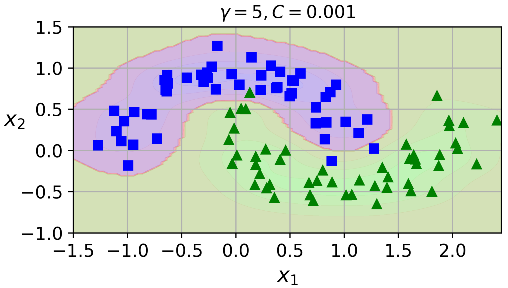


Radiális bázis függvények kernelként

Ahogy a polinomikus jellemzők esetében, úgy **a hasonlósági jellemzők esetén is működik a kernel trükk**.

A diagramon különböző γ és C értékekkel tanított modellek láthatóak.

A γ növelése szűkebb haranggörbékkel végzi az illesztést, ezért minden egyed hasonlósági tartománya kisebb. A γ csökkentése nagyobb befolyási teret ad a pontoknak.

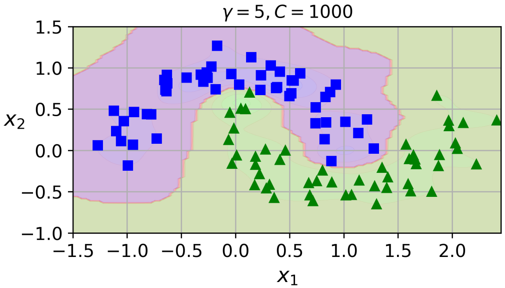


Radiális bázis függvények kernelként

Ahogy a polinomikus jellemzők esetében, úgy **a hasonlósági jellemzők esetén is működik a kernel trükk**.

A diagramon különböző γ és C értékekkel tanított modellek láthatóak.

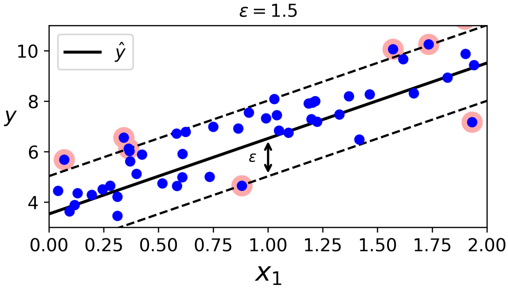
A γ növelése szűkebb haranggörbékkel végzi az illesztést, ezért minden egyed hasonlósági tartománya kisebb. A γ csökkentése nagyobb befolyási teret ad a pontoknak.



SVM regresszió

Az SVM regresszió esetében a cél ellentétes az osztályozáséval. A legnagyobb út helyett az algoritmus megpróbálja **a lehető legtöbb egyedet az útra illeszteni a margósértések minimumon tartása mellett**.

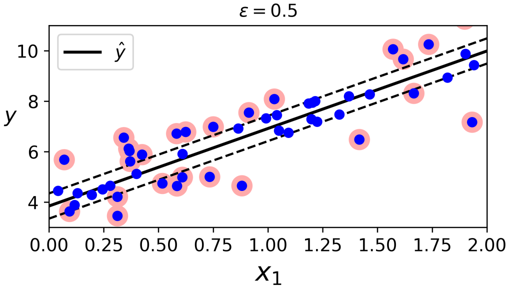
Az út szélességét a ε paraméter szabályozza. Nagyon ε paraméter nagyobb margót jelent. A margón belül hozzáadott egyedek nem befolyásolják a modell definíciót. Ez az ε-érzéketlenség.



SVM regresszió

Az SVM regresszió esetében a cél ellentétes az osztályozáséval. A legnagyobb út helyett az algoritmus megpróbálja **a lehető legtöbb egyedet az útra illeszteni a margósértések minimumon tartása mellett**.

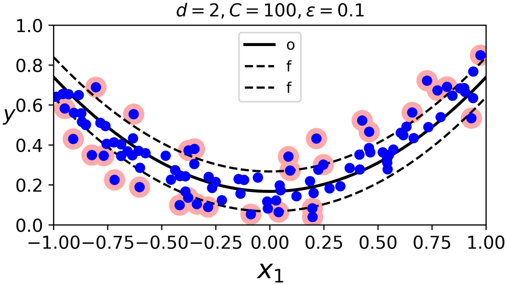
Az út szélességét a ε paraméter szabályozza. Nagyon ε paraméter nagyobb margót jelent. A margón belül hozzáadott egyedek nem befolyásolják a modell definíciót. Ez az ε-érzéketlenség.



Nemlineáris SVM regresszió

Kernelizált SVM regresszorok használata lehetséges regressziós problémák esetén is.

A következő diagramokon másodfokú polinomikus kernellel tanított SVM regresszorok láthatóak.



Nemlineáris SVM regresszió

Kernelizált SVM regresszorok használata lehetséges regressziós problémák esetén is.

A következő diagramokon másodfokú polinomikus kernellel tanított SVM regresszorok láthatóak.