Osztályozási módszerek

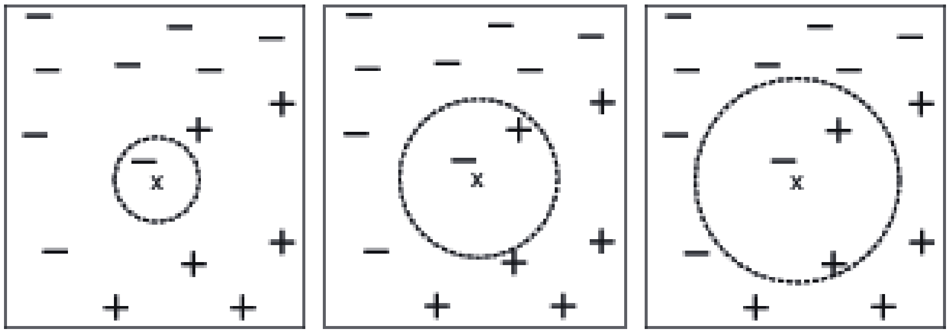
Legközelebbi szomszéd osztályozók

- Az **osztályozás** egy kétlépcsős eljárást foglal magába:
- induktív lépés: az adatokból egy osztályozási modellt alkotunk
- deduktív lépés: a modell tesztesetekre való alkalmazásához
- **Rote osztályozó**: az összes tanulóadatot memorizálja, és csak akkor osztályoz, ha a tesztpéldány attribútumai pontosan illeszkednek egy tanulóesetre.
- A módszer egy nyilvánvaló hátulütője az, hogy bizonyos tesztesetek nem osztályozhatók, mivel nem egyeznek meg a tanulóesetek egyikével sem.

Legközelebbi szomszéd osztályozók

- A módszer rugalmasabbá tétele: a teszteset attribútumaihoz viszonylag hasonló valamennyi tanulóeset megkeresése. Ezek az esetek a **legközelebbi szomszédok (nearest neighbors)**, felhasználhatók a teszteset osztálycímkéjének meghatározásához.
- A legközelebbi szomszédok használatának indoklását legjobban a következő mondás szemlélteti: ''Ha valami úgy totyog, mint egy kacsa, úgy hápog, mint egy kacsa és úgy néz ki, mint egy kacsa, akkor az valószínűleg egy kacsa. ''
- A legközelebbi szomszéd osztályozó minden egyes esetet egy adatpontként reprezentál egy d-dimenziós térben, ahol d az attribútumok száma. Szomszédsági mértékek valamelyikével kiszámítjuk ezek közelségét a tanulóhalmaz összes többi adatpontjához.

Legközelebbi szomszéd osztályozók



(a) 1-legközelebbi szomszéd (b) 2-legközelebbi szomszéd (c) 3-legközelebbi szomszéd

- A többségi szavazási sémával az adatpontot a pozitív osztályhoz rendeljük hozzá.
- Holtverseny esetén az adatpont osztályozásához véletlenszerűen választhatjuk valamelyik osztályt.

Legközelebbi szomszéd osztályozók

- A k érték helyes megválasztásának fontossága:
- Ha k túl kicsi, akkor a legközelebbi szomszéd osztályozó a tanulóadatokban jelenlevő zaj miatt hajlamos lehet a túlillesztésre.
- Ha k túl nagy, akkor a legközelebbi szomszéd osztályozó rosszul osztályozhatja a tesztpéldányt, mivel a legközelebbi szomszédok listája a szomszédságtól messzi adatpontokat is tartalmazhat.
- Az algoritmus kiszámítja minden teszteset és tanulóeset távolságát (hasonlóságát) -> számításköltséges nagy-számú tanulóeset esetén -> hatékony indexelési eljárások, amelyek csökkentik az egyes tesztesetek legközelebbi szomszédainak megkereséséhez szükséges számítási mennyiséget.

Legközelebbi szomszéd osztályozók

- **példányalapú tanulás:** konkrét tanulópéldányokat használ predikció végzéséhez anélkül, hogy az adatokból származó absztrakcióra (modellre) lenne szüksége.
- egy szomszédsági mérték szükséges a példányok hasonlóságának vagy távolságának meghatározásához, valamint egy osztályozási függvény, amely más példányokhoz közelsége alapján visszaadja egy tesztpéldány előrejelzett osztályát.
- Nem igényelnek modellépítést, viszont a tesztesetek osztályozása elég költséges lehet, mivel külön-külön kell kiszámolnunk a teszt- és a tanulóesetek közelségét.

Legközelebbi szomszéd osztályozók

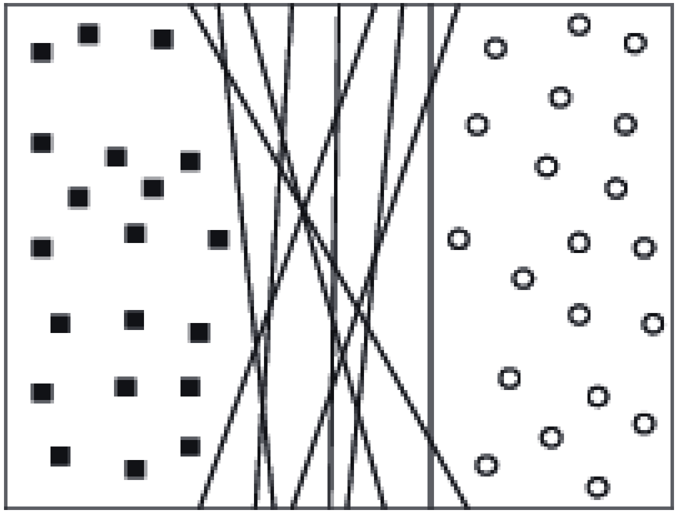
- lokális információk alapján végeznek előrejelzést. A döntéshozatal lokálisan történik a osztályozás során, kis k értékek esetén ezek az osztályozók elég érzékenyek a zajra.
- hibás előrejelzéseket adhatnak, hacsak nem végezzük el a megfelelő előfeldolgozási lépéseket a szomszédsági mértéken és az adatokon.
- magasság (méter) és testsúly (font) alapján osztályozás. A magasság attribútum kis változékonyságú (1,5 és 1,85 között), míg a testsúly attribútum 90 és 250 font között változhat –> az attribútumok skáláját figyelembe kell venni

Tartóvektor-gép (SVM)

- Egy jelentős figyelmet kiváltó osztályozási módszer a tartóvektor-gép (SVM - Support Vector Machine).
- A módszer a statisztikai tanulás elméletéből származik és ígéretes tapasztalati eredményeket mutat sok gyakorlati alkalmazásban, a kézzel írt számjegyek felismerésétől kezdve a szövegosztályozásig.
- Az SVM nagyon jól működik sokdimenziós adatokkal, elkerüli a dimenzióproblémát.
- A módszer egy másik egyedi jellege, hogy a döntési határt a tanulóesetek egy részhalmazának segítségével reprezentálja -> tartóvektorok (support vector)

Tartóvektor-gép (SVM)

- Az SVM alapötletének szemléltetése: bevezetjük a **maximális margójú hipersík (maximal margin hyperplane)** fogalmát és megmagyarázzuk egy ilyen hipersík választásának értelmét.



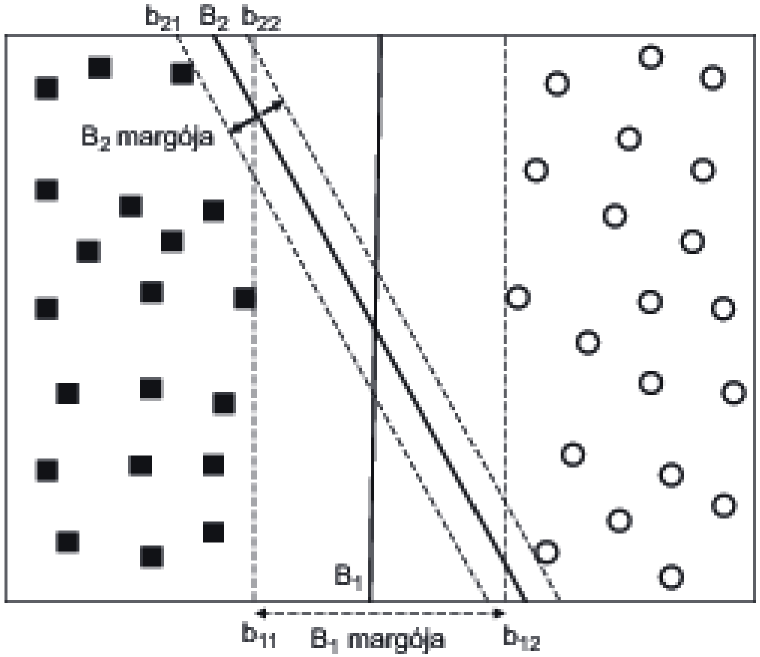
- két különböző osztályba tartozó eseteket tartalmaz

Tartóvektor-gép (SVM)

- Az adathalmaz lineárisan szeparálható, azaz találhatunk olyan hipersíkot, amelynek egyik oldalán az összes négyzet, a másik oldalán az összes kör van.
- Végtelen sok ilyen hipersík lehetséges.
- Noha ezeknek a tanulóhalmazon mért hibája nulla, nincs garancia arra, hogy a hipersíkok egyformán jól fognak teljesíteni korábban még nem látott esetekre.
- Az osztályozó ki kell, hogy válassza a hipersíkok egyikét a döntési határ reprezentálásához annak alapján, hogy várhatólag milyen jól teljesítenek a teszteseteken.

Tartóvektor-gép (SVM)

- a hipersíkok különböző megválasztása milyen hatással van az általánosítási hibára?



- Tekintsünk két döntési határt, B_1-et és B_2-t.

Tartóvektor-gép (SVM)

- Mindkét döntési határ szét tudja választani a tanulóeseteket a megfelelő osztályokra, félreosztályozási hiba elkövetése nélkül.
- Mindegyik B_i döntési határhoz tartozik két hipersík (b_{i1} és b_{i2}). b_{i1} -et úgy kapjuk meg, hogy addig tolunk el a döntési határtól egy vele párhuzamos hipersíkot, amíg az nem érinti a legközelebbi négyzetet, míg b_{i2} esetént az eltolt hipersík nem érinti a legközelebbi kört.
- A két hipersík közötti távolságot nevezzük az **osztályozó margójának**.
- B_1 margója jelentősen nagyobb, mint B_2-é. A példában B_1 bizonyul a tanulópéldányok maximális margójú hipersíkjának.

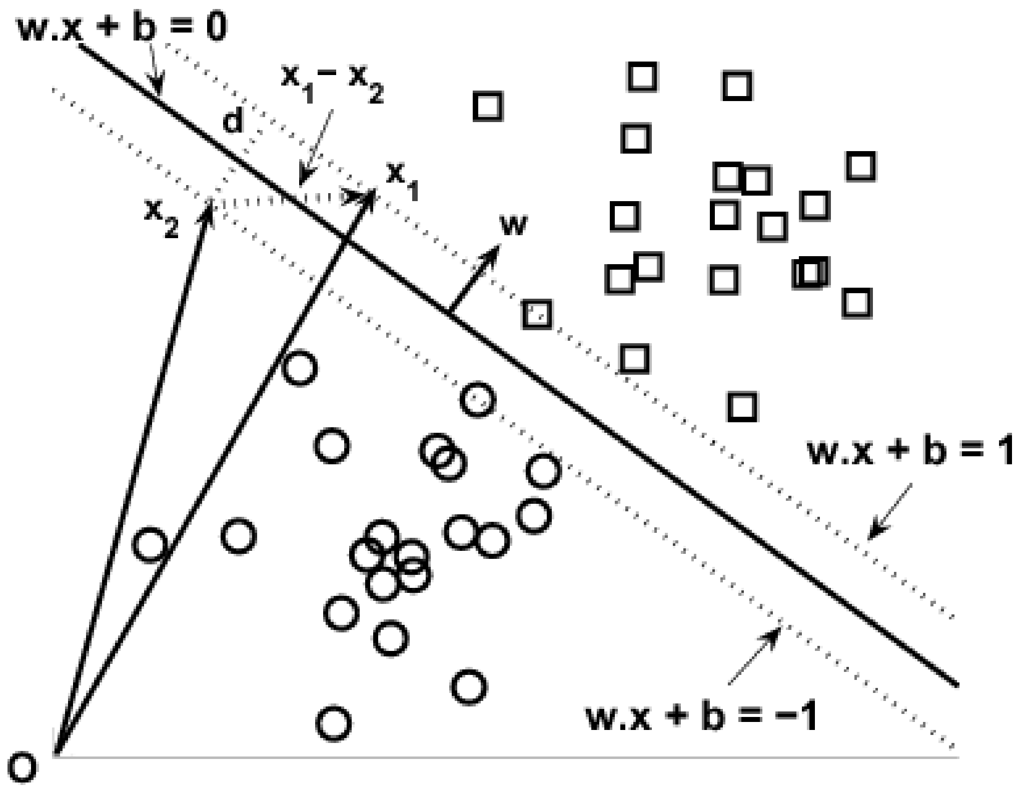
A maximális margó indoklása

- A nagy margóval rendelkező döntési határoknak általában jobb az általánosítási hibájuk, mint a kis margójúaké. Ha a margó kicsi, akkor a döntési határ bármilyen kis perturbációjának elég jelentős hatása lehet az osztályozásra. A kis margóval rendelkező döntési határokat létrehozó osztályozók ezért hajlamosabbak a modell túlillesztésre és korábban nem látott eseteken gyakran rosszul általánosítanak.
- Egy formálisabb magyarázatot ad a lineáris osztályozó margójának és általánosítási hibájának kapcsolatára a strukturális kockázat minimalizálásként (SRM) ismert statisztikai tanulási elv. Ez az elv egy felső korlátot ad egy osztályozó általánosítási hibájára.

Lineáris SVM: szeparálható eset

- A lineáris SVM egy olyan osztályozó, amely egy a legnagyobb margóval rendelkező hipersíkot keres -> maximális margójú osztályozó
- **Lineáris döntési határ**
- egy N számú tanulóesetből álló bináris osztályozási probléma: Mindegyik esetet egy (x_i , y_i) n-es jelöli, ahol $x_i=(x_{i1}, x_{i2},\dots, x_{id})^T$ az i. eset attribútumhalmaza. Megállapodás szerint: $y_i \in \{-1,1\}$ az osztálycímke.
- A lineáris osztályozó döntési határa:
- **w·x+b=0**, ahol w és b a modell paraméterei.

Lineáris SVM: szeparálható eset



- egy négyzetekből és körökből álló kétdimenziós tanulóhalmaz a tanulóeseteket a megfelelő osztályokra kettéválasztó döntési határral

Lineáris SVM: szeparálható eset

- Ha +1 osztályúnak címkézünk minden négyzetet és -1 osztályúnak minden kört, akkor az alábbi módon tudjuk prediktálni bármely teszteset osztálycímkéjét:

$$y = \begin{cases} 1, & ha\ w \cdot z + b \geq 0, \\ -1, & ha\ w \cdot z + b \leq 0. \end{cases}$$

- **A lineáris osztályozó margója**
- Tekintsük a döntési határhoz legközelebbi négyzetet és kört. Átskálázhatjuk úgy a döntési határ *w* és *b* paramétereit, hogy a két párhuzamos hipersík:

$$b_{i1}: w \cdot x + b = 1,$$

$$b_{i2}: w \cdot x + b = -1.$$

Lineáris SVM: szeparálható eset

- A döntési határ *d* margóját a két hipersík közötti távolság adja meg, amely meghatározható a két egyenlet különbségének segítségével:

$$d = \frac{2}{\|w\|}.$$

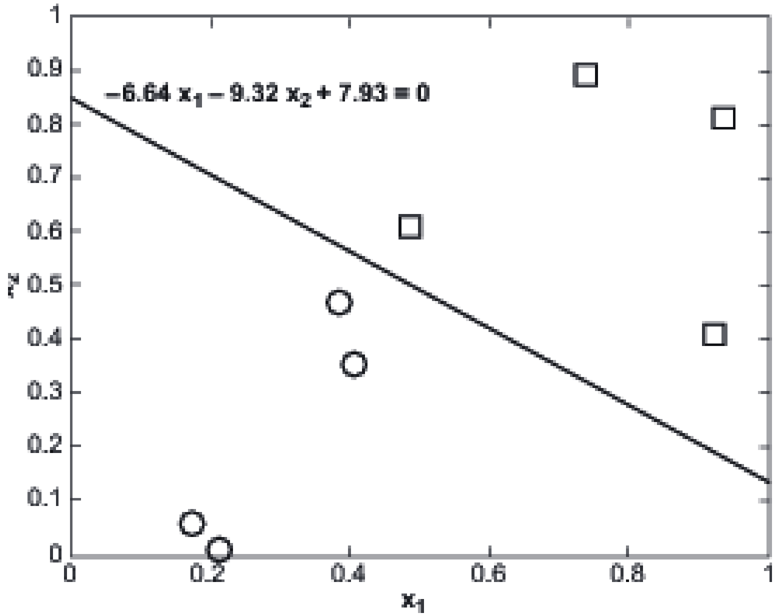
- **A lineáris SVM modell tanulása**
- Az SVM tanulási fázisa magában foglalja a döntési határ *w* és *b* paramétereinek a tanulóadatokból történő megbecslését. A paramétereket úgy kell megválasztani, hogy az alábbi feltétel teljesüljön:
- $y_i\ (w \cdot x_i + b) \geq 1, \quad (i=1,2,\ldots,N)$

Lineáris SVM: szeparálható eset

- Bár az előző feltételek alkalmazhatók bármely lineáris osztályozóra, az SVM egy további követelményt támaszt: a döntési határ margója maximális legyen. A margó maximalizálása azonban ekvivalens az alábbi célfüggvény minimalizálásával:
- $f(w) = \| w \|^2 / 2$.
- Mivel az f(w) célfüggvény kvadratikus és a korlátozások lineárisak a *w* és *b* paraméterekben, ezt **konvex optimalizálás**nak nevezzük, amely a szokásos Lagrange-multiplikátor módszerrel oldható meg.

Lineáris SVM: szeparálható eset

- Példa: kétdimenziós adathalmaz, nyolc tanulópéldány



- Kvadratikus programozás segítségével megoldhatjuk az optimalizálási problémát, hogy minden tanulópéldányra megkapjuk a Lagrange-multiplikátort.

Lineáris SVM: szeparálható eset

- A Lagrange-multiplikátorok a táblázat utolsó oszlopában kerülnek leírásra. Vegyük észre, hogy csak az első két példánynak van nullától különböző Lagrange-multiplikátora. Ezek a példányok felelnek meg az adathalmaz tartóvektorainak:

x_1	x_2	y	Lagrange multiplikátor
0.3858	0.4687	1	65.5261
0.4871	0.611	–1	65.5261
0.9218	0.4103	–1	0
0.7382	0.8936	–1	0
0.1763	0.0579	1	0
0.4057	0.3529	1	0
0.9355	0.8132	–1	0
0.2146	0.0099	1	0

Lineáris SVM: szeparálható eset

- $w=(w_1, w_2)$ és b a döntési határ paraméterei:

$$w_1 = \sum_i \lambda_i y_i x_{i1} = 65{,}5261 \times 1 \times 0{,}3858 + 65{,}5261 \times -1 \times 0{,}4871 = -6{,}64.$$

$$w_2 = \sum_i \lambda_i y_i x_{i2} = 65{,}5261 \times 1 \times 0{,}4687 + 65{,}5261 \times -1 \times 0{,}611 = -9{,}32.$$

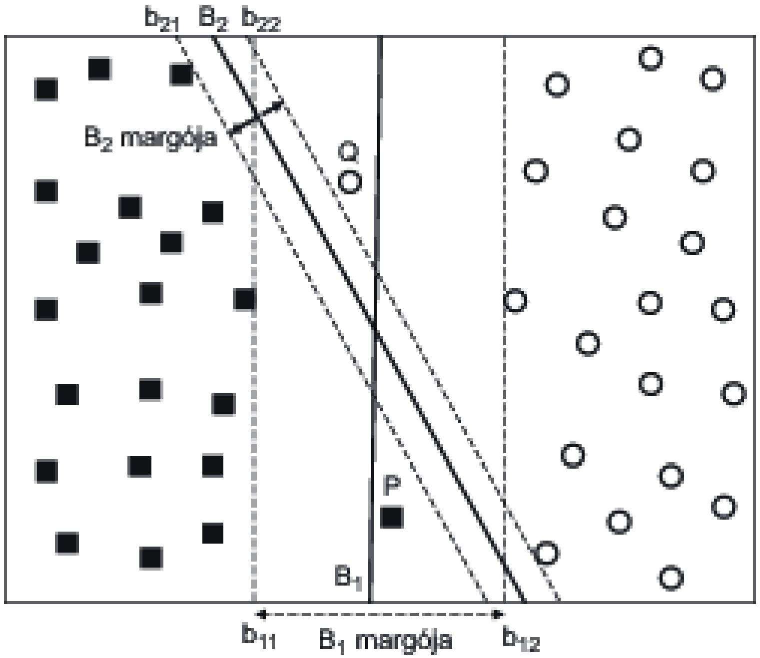
- A b torzítás tag az egyes tartóvektorokra:

$$b^{(1)} = 1 - w \cdot x_1 = 1 - (-6{,}64)(0{,}3858) - (-9{,}32)(0{,}4687) = 7{,}9300.$$

$$b^{(2)} = -1 - w \cdot x_2 = -1 - (-6{,}64)(0{,}4871) - (-9{,}32)(0{,}611) = 7{,}9289.$$

- Ezeket átlagolva azt kapjuk, hogy b=7,93.

Lineáris SVM: nem szeparálható eset



- két új eset (P és Q)
- Noha a B_1 döntési határ hibásan osztályozza az új eseteket, míg B_2 helyesen, ez nem jelenti azt, hogy B_2 jobb döntési határ, mint B_1, mert az új esetek a tanulóadatokban lévő zajnak felelhetnek meg.

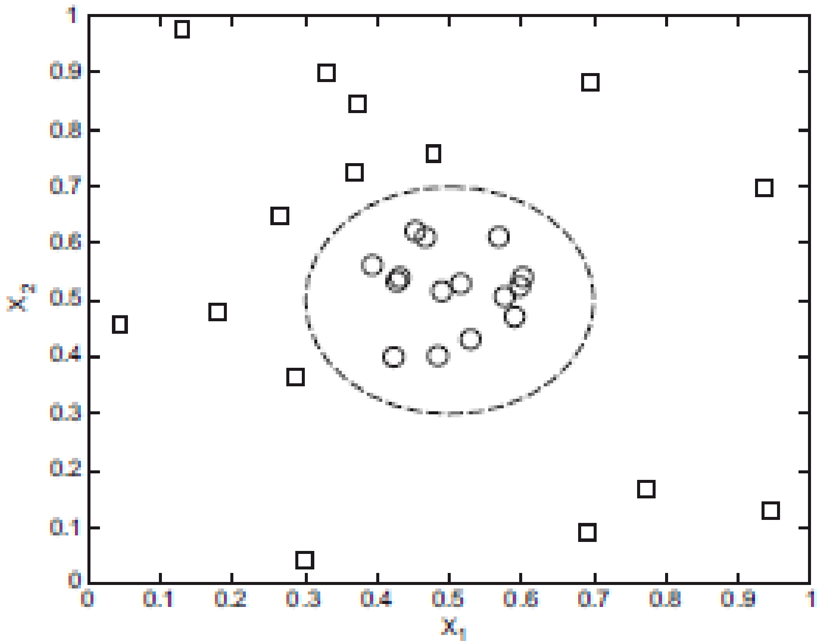
Lineáris SVM: nem szeparálható eset

- Továbbra is B_1-et kell előnyben részesíteni B_2-vel szemben, mert szélesebb a margója, így kevésbé hajlamos a túlillesztésre. Az eddigi SVM azonban csak olyan döntési határokat hozott létre, amelyek hibamentesek.
- hogyan módosítható a **puha margóként (soft margin)** ismert módszer használatával egy olyan döntési határ megtanulása: tolerálja a kis hibákat a tanulóhalmazon.
- még olyan helyzetekben is lehetővé teszi az SVM számára lineáris döntési határ létrehozását, ahol az osztályok lineárisan nem szeparálhatók. Ennek érdekében az SVM tanuló algoritmusa figyelembe veszi a margó szélessége és a lineáris döntési határ által a tanulóhalmazon elkövetett hibák száma közötti kompromisszumot.

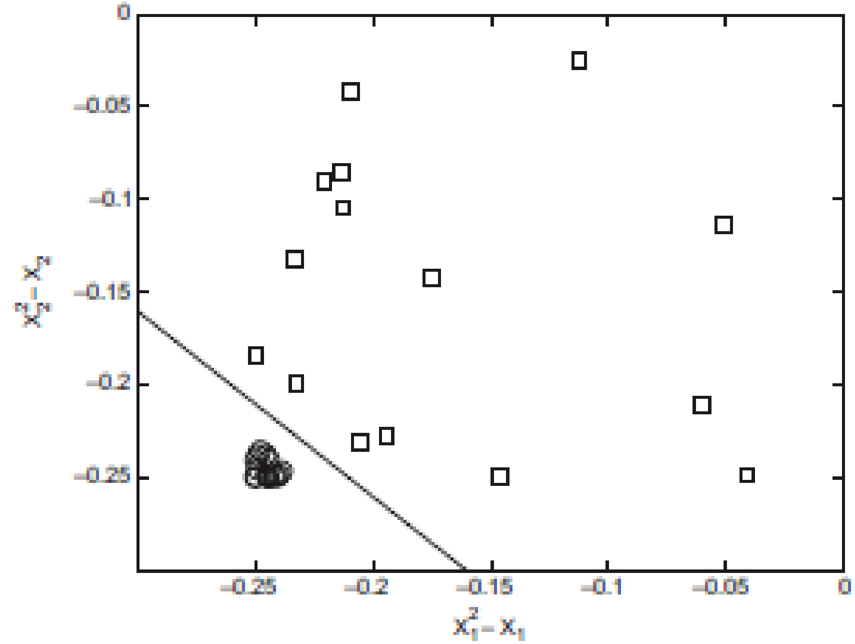
Nemineáris SVM

- Az SVM korábbi megfogalmazásai egy lináris döntési határt hoztak létre a tanulóesetek osztályoknak megfelelő szétválasztásához.
- Újabb módszertan: az SVM olyan adatokra való alkalmazásához, amelyek döntési határa nemlineáris.
- A trükk itt az adatok áttranszformálása az eredeti koordinátatérből egy új térbe úgy, hogy a transzfor-mált térben a példányok egy lineáris döntési határral legyenek szétválaszhatók.
- A transzformáció elvégzése után alkalmazhatjuk az előző módszertant ahhoz, hogy egy lineáris döntési határt találjunk a transzformált térben.

Nemineáris SVM



(a) Döntési határ az eredeti kétdimenziós térben



(b) Döntési határ a transzformált térben

Bayes-féle osztályozók

- Az attribútumhalmaz és az osztályváltozó közötti kapcsolat sok alkalmazásban nemdeterminisztikus.
- Egy tesztrekord osztálycímkéje nem jelezhető előre teljes bizonyossággal még akkor sem, ha az attribútumhalmaza azonos valamelyik tanulóesettel.
- Ez a helyzet a zajos adatok vagy bizonyos zavaró, az osztályozást befolyásoló, de az elemzésben nem szereplő hatások jelenléte miatt merülhet fel.
- Egészségügyi kockázat meghatározása a személy étrendje és edzéseinek gyakorisága alapján.
- más tényezők: például az öröklődés, a mértéktelen dohányzás és túlzott alkoholfogyasztás.

Bayes-féle osztályozók

- **Bayes-tétel:** egy statisztikai elv az osztályokra vonatkozó a priori tudásnak az adatokból gyűjtött új tényekkel történő kombinálására.
- Legyen X és Y véletlen változók. A P(X=x, Y=y) együttes valószínűség annak valószínűségét jelenti, hogy az X változó az x és az Y változó az y értéket veszi fel.
- P(Y=y|X=x): feltételes valószínűsége annak, hogy az Y válto-zó az y értéket veszi fel, feltéve, hogy X megfigyelt értéke x.
- X és Y együttes és feltételes valószínűsége között az alábbi összefüggés áll fenn: P(X,Y)=P(Y|X) P(X)=P(X|Y) P(Y).
- Átrendezés után kapjuk meg a Bayes-tételt:
- **P(Y|X)= P(X|Y) P(Y) / P(X)**

A Bayes-tétel felhasználása osztályozásra

- Jelölje X az attribútumhalmazt és Y az osztályváltozót. Ha az osztályváltozónak nemdeterminisztikus kapcsolata van az attribútumokkal, akkor X és Y véletlen változóként kezelhető, kapcsolatuk valószínűségileg P(Y|X) segítségével írható le. Ez a feltételes valószínűség az Y **a posteriori** valószínűsége, a P(Y) **a priori** valószínűségével szemben.
- A tanítási fázis során a P(Y|X) a posteriori valószínűségeket X és Y minden kombinációjára meg kell tanulnunk a tanulóadatokból szerzett információk alapján.
- Ezen valószínűségek ismeretében egy X' tesztrekord úgy osztályozható, hogy megkeressük a P(Y'|X') a posteriori valószínűséget maximalizáló Y' osztályt.

A Bayes-tétel felhasználása osztályozásra

- egy hitelfelvevő késedelmesen fog-e fizetni? előrejelzésének feladata

	bináris	kategorikus	folytonos	osztály
Tra.	Háztulaj-donos	Családi állapot	Éves jö-vedelem	Késedelmes hitelfelvevő
1	Igen	egyedülálló	125 000	Nem
2	Nem	házas	100 000	Nem
3	Nem	egyedülálló	70 000	Nem
4	Igen	házas	120 000	Nem
5	Nem	elvált	95 000	Igen
6	Nem	házas	60 000	Nem
7	Igen	elvált	220 000	Nem
8	Nem	egyedülálló	85 000	Igen
9	Nem	házas	75 000	Nem
10	Nem	egyedülálló	90 000	Igen

Tra. = Tranzakcióazonosító

A Bayes-tétel felhasználása osztályozásra

- tanulóhalmaz a következő attribútumokkal: Lakástulajdonos, Családi állapot és Éves jövedelem.
- A fizetési késedelembe esett hitelfelvevőket Igen-ként osztályozzuk, míg Nem-ként azokat, akik visszafizették a kölcsönt.
- Adott egy tesztrekord a következő attribútumhalmazzal: X=(Lakástulajdonos=Nem, Családi állapot=Házas, Éves jövedelem = 120000). A rekord osztályozásához ki kell számolnunk a P(Igen|X) és P(Nem|X) a posteriori valószínűségeket a tanulóadatok alapján.
- Ha P(Igen|X)≥P(Nem|X), akkor a rekordot Igen-ként osztályozzuk, egyébként Nem-ként.

A Bayes-tétel felhasználása osztályozásra

- Bonyolult probléma pontosan becsülni az a posteriori valószínűségeket az osztálycímke és attribútumérték összes lehetséges kombinációjára, mivel nagyon nagy tanulóhalmazt igényel, még csekély számú attribútumra is. A Bayes-tétel azért hasznos, mert lehetővé teszi a P(Y|X) a posteriori valószínűség kifejezését.
- Különböző Y értékek a posteriori valószínűségeinek összehasonlításánál a P(X) nevező tag mindig állandó, így figyelmen kívül hagyható.
- A P(Y) a priori valószínűség könnyen becsülhető a tanulóhalmazból az egyes osztályokhoz tartozó tanulórekordok arányának kiszámításával.
- A P(X|Y) osztályra vonatkozó feltételes valószínűségek becsléséhez: naiv Bayes-féle osztályozó

Naiv Bayes osztályozó

- A naiv Bayes-féle osztályozó az osztályra vonatkozó feltételes valószínűséget az alapján becsüli meg, hogy az attribútumok feltételesen függetlenek adott y osztálycímke mellett.
- **Feltételes függetlenség:** P(X|Y=y) = $\prod_{\mathrm{i}=1}^{\mathrm{d}} \mathrm{P}(\mathrm{X}_i|\mathrm{Y}=\mathrm{y})$, ahol minden X = {$X_1$, X_2,…, X_d} attribútumhalmaz d attribútumból áll.
- Jelöljék X, Y és Z valószínűségi változók három halmazát. Az X-beli változókat feltételesen függetleneknek mondjuk Y-tól Z mellett, ha teljesül a következő feltétel:
- P(X|Y, Z)=P(X|Z).

Feltételes függetlenség

- A feltételes függetlenség egy példája a karhossz és az olvasási készségek közötti kapcsolat.
- Úgy tapasztalhatjuk, hogy a hosszabb karú emberek többnyire magasabb szintű olvasási készségekkel rendelkeznek.
- Ez a kapcsolat egy zavaró tényező jelenlétével magyarázható, amely az életkor. Egy kisgyermeknek általában rövid karjai vannak és nem állnak rendelkezésére egy felnőtt olvasási készségei.
- Ha egy személy életkora rögzített: eltűnik a karhossz és az olvasási készségek között megfigyelt kapcsolat. Megállapíthatjuk így, hogy a karhossz és az olvasási készségek feltételesen függetlenek, ha az életkor változó rögzített.

Naiv Bayes osztályozó

- Feltételes függetlenség esetén az osztályra vonatkozó feltételes valószínűség X minden kombinációjához történő kiszámítása helyett csupán az egyes X_i-k feltételes valószínűségét kell megbecsülnünk adott Y mellett.
- Ez a megközelítés gyakorlati szempontból sokkal alkalmasabb, mivel nem igényel nagyon nagy tanulóhalmazt ahhoz, hogy jó becslést kapjuk a valószínűségre.
- Tesztrekord osztályozásához a naiv Bayes-osztályozó minden egyes Y osztály a posteriori valószínűségét számítja ki:
- $P(Y|X) = P(Y) \prod_{\mathrm{i}=1}^{\mathrm{d}} P(X_i|Y = y) \,/\, P(X)$.
- Mivel P(X) rögzített minden Y-ra, elegendő a számlálót, $P(Y)\prod_{\mathrm{i}=1}^{\mathrm{d}} P(X_i|Y = y)$ maximalizáló osztályt választani.

Naiv Bayes osztályozó

- **Kategorikus attribútumok feltételes valószínűségeinek becslése**
- Egy X_i kategorikus attribútum $P(X_i = x_i | Y=y)$ feltételes való-színűségének becslése az y osztályban egy bizonyos x_i attri-bútumértéket felvevő tanulópéldányok aránya szerint történik.
- A korábbi példában adott tanulóhalmazban a hét hitelét visszafizető ember közül háromnak van saját lakása is. Ennek következtében a P(Lakástulajdonos = Igen|Nem) feltételes valószínűség 3/7-del egyenlő. Hasonlóképpen P(Családi állapot=Egyedülálló|Igen)=2/3 adja meg azoknak a nem fizető adósoknak a feltételes valószínűségét, akik egyedülállóak.

Naiv Bayes osztályozó - Folytonos attribútumok

- Két mód van folytonos attribútumok osztályra vonatkozó feltételes valószínűségeinek becslésére:

1. Diszkretizálhatunk minden folytonos tulajdonságot, majd ezt követően a megfelelő diszkrét intervallummal helyettesíthetjük a folytonos attribútumértékeket.

- A módszer ordinális attribútumokká alakítja a folytonos attribútumokat.
- A $P(X_i|Y=y)$ feltételes valószínűség becslése azoknak az y osztályba tartozó tanulórekordoknak az arányának kiszámításával történik, amelyek a megfelelő X_i intervallumba esnek.

Naiv Bayes osztályozó - Folytonos attribútumok

- A becslési hiba a diszkretizálási stratégiától, valamint a diszkrét intervallumok számától függ.
- Ha túl nagy az intervallumok száma, túl kevés tanulórekord van az egyes intervallumokban $P(X_i|Y)$-ra megbízható becslés biztosításához.
- Ha túl kicsi az intervallumok száma, akkor bizonyos intervallumok különböző osztályokba tartozó rekordokat aggregálhatnak, és elhibázhatjuk a helyes döntési határt.

2. A folytonos változóra feltételezhetünk egy bizonyos fajta valószínűségi eloszlást, és a tanulóadatok segítségével becsülhetjük meg az eloszlás paramétereit.

- A normális eloszlást gyakran választják folytonos attribú-tumok osztályra vonatkozó feltételes valószínűségének reprezentálásához.

Példa a naiv Bayes-féle osztályozóra

Az előző példában minden egyes kategorikus attribútumra kiszámíthatjuk az osztályra vonatkozó feltételes valószínűséget a folytonos attribútum mintaátlagával és varianciájával együtt.

Tid	Lakástu-lajdonos	Családi állapot	Éves jö-vedelem	Nem fizető adós
1	Igen	Egyedülálló	125 000	Nem
2	Nem	Házas	100 000	Nem
3	Nem	Egyedülálló	70 000	Nem
4	Igen	Házas	120 000	Nem
5	Nem	Elvált	95 000	Igen
6	Nem	Házas	60 000	Nem
7	Igen	Elvált	220 000	Nem
8	Nem	Egyedülálló	85 000	Igen
9	Nem	Házas	75 000	Nem
10	Nem	Egyedülálló	90 000	Igen

(a)

P(Lakástulajdonos=Igen|Nem) = 3/7
P(Lakástulajdonos=Nem|Nem) = 4/7
P(Lakástulajdonos=Igen|Igen) = 0
P(Lakástulajdonos=Nem|Igen) = 1
P(Családi állapot=Egyedülálló|Nem) = 2/7
P(Családi állapot=Elvált|Nem) = 1/7
P(Családi állapot=Házas|Nem) = 4/7
P(Családi állapot=Egyedülálló|Igen) = 2/3
P(Családi állapot=Elvált|Igen) = 1/3
P(Családi állapot=Házas|Igen) = 0

Az éves jövedelemre:
Ha osztály=Nem: mintaátlag=110
variancia=2975
Ha osztály=Igen: mintaátlag=90
variancia=25

(b)

Példa a naiv Bayes-féle osztályozóra

- Az X=(Lakástulajdonos=Nem, Családi állapot=Házas, Jövedelem=120) tesztrekord osztálycímkéjének előrejelzéséhez a P(Nem|X) és P(Igen|X) a posteriori valószínűségek kiszámítása szükséges.
- Ezekre az a posteriori valószínűségekre becslés adható, ha kiszámoljuk a P(Y) a priori valószínűség és a $\prod_{i=1}^{d} P(X_i|Y=y)$ osztályra vonatkozó feltételes valószínűségek szorzatát.
- Az egyes osztályok a priori valószínűsége megbecsülhető, ha minden osztályra kiszámoljuk az odatartozó tanulórekordok arányát. Mivel három Igen osztályba tartozó rekord és hét Nem osztályba tartozó rekord van, P(Igen)=0,3 és P(Nem)=0,7 .

Példa a naiv Bayes-féle osztályozóra

- Az osztályra vonatkozó feltételes valószínűségek:
- P(X|Nem)=P(Lakástulajdonos = Nem|Nem) P(Állapot = Házas|Nem) P(Éves jövedelem = 120|Nem) = $4/7 \times 4/7 \times 0{,}0072 = 0{,}0024$.
- P(X|Igen)=P(Lakástulajdonos = Nem|Igen) P(Állapot = Házas|Igen) P(Éves jövedelem = 120|Igen) = $1 \times 0 \times 1{,}2 \times 10^{-9} = 0$.
- A Nem osztály a posteriori valószínűsége P(Nem|X) = $\alpha \times 7/10 \times 0{,}0024 = 0{,}0016\alpha$ ($\alpha = 1/P(X)$ konstans tag). Hasonló módszerrel mutathatjuk meg azt, hogy az Igen osztály a posteriori valószínűsége nulla.
- Mivel P(Nem|X)≥P(Igen|X), a rekordot Nem-ként osztályozzuk.

Bevezetés az adatbányászatba (Pang-Ning Tan, Michael Steinbach, Vipin Kumar) című tananyaga alapján készült (részben)