OTDK-dolgozat

2019

Kujbus Marcell

Budapesti Corvinus Egyetem

Közgazdaságtudományi Kar

VÁLLALATI FELVÁSÁRLÁS SIKERESSÉGÉNEK ELŐREJELZÉSE SUPPORT VECTOR MACHINE-NEL

FORECASTING THE OUTCOME OF A CORPRATIONAL MERGER AND ACQUISITON DEAL VIA SUPPORT VECTOR MACHINE

Témavezető

Dr. Ferenci Tamás

Kézirat lezárásának dátuma: Budapest, 2018.11.25

Tartalomjegyzék

Táblázatok jegyzéke

Ábrák jegyzéke

1. fejezet

Bevezetés

Egy vállalati felvásárlásnak (Merger&Acquisition - M&A ügyletnek) a bejelentése előtt történő előrejelzése már számos szerző érdeklődését felkeltette, ami talán annak is köszönhető, hogy egy valóban használható előrejelző modell szignifikáns kockázattal korrigált hozamot biztosítana. Ahogy azt a korábbi szakirodalmak bemutatásánál ismertetem, az eredmények vegyes képet mutatnak: egyes szerzők szignifikánsnak tekintik a modelljük által a teszthalmazon elért pontossági mutató értékét, míg más szerzők konklúziója, hogy statisztikai modelljük a puszta véletlen találgatásnál nem tud jobb eredményt produkálni.

Egy vállalati felvásárlás elméletileg akkor lehetséges, ha a két vállalat együttes értéke nagyobb, mint azok értéke külön-külön, így a felvásárló vállalat hajlandó a felvásárolandó vállalat jelenlegi piaci kapitalizációjához képest prémiumot fizetni a tranzakció létrejötte érdekében. A felvásárló vállalat maximális fizetési hajlandósága pedig a szinergiákból vagy más forrásból keletkező többletérték. A fizetett prémiummal arányosan így növekedni fog a felvásárlandó vállalat piaci értéke, amennyiben a felvásárlás sikeres lesz. Minél nagyobb a felvásárló vállalatok közötti verseny, várhatóan annál közelebb lesz a felvásárlandó vállalat részvényárfolyamának emelkedése a felvásárlás hatására keletkező többletértékhez. Ha sikerülne pozitív kockázattal korrigált többlethozammal rendelkező stratégiát konstruálni egy statisztikai modellre alapozva, az ellentmondana a hatékony piacok elmélete közepesen erős változatának Bodie *et al.* (2013) alapján, amennyiben a nyilvánosan elérhető pénzügyi adatokat (mint például a ROE – saját tőke arányos megtérülés) fundamentális információnak tekintjük. Így ha a dolgozatomban prezentált modell szignifikáns eredményt is produkál, az eredményekre tekintsünk végig szkeptikusan, és fordítsunk különös figyelmet azokra a feltevésekre illetve limitációkra, amik szerepet játszhatnak az eredmények alakulásában.

A vállalati felvásárlások sikerességének előrejelzésére vonatkozó korábbi munkák elsősorban két módszerre koncentrálódnak: a lineáris diszkriminancia-analízisre és a logisztikus regresszióra. Szakdolgozatomban az ezen a területen kevésbé elterjedt, de ígéretesnek tűnő gépi tanulási módszert alkalmazok, nevezetesen a Support Vector Machine-t (SVM). Az SVM módszer kevésbé elterjedt, így annak elméleti hátterét részletesen is bemutatom. A modellbe bevont változókat a vállalati felvásárlások szakirodalma alapján válogattam össze, és a Bloomberg MA valamint EQS függvényei segítségével töltöttem le azokat. Igyekeztem jelentős mennyiségű, továbbá használható rekordokból álló adatbázist felépíteni. A megfigyelések száma így (23 450 fel nem vásárolt és 2 551 felvásárolt vállalat) különösen magas a korábbi munkákhoz képest (legfeljebb 1 100-1 200 megfigyelés), ami a Budapesti Corvinus Egyetem Finlab termében rendelkezésre álló Bloomberg terminálnak köszönhető. A modellt teljes egészében az R Studio statisztikai programcsomag segítségével implementáltam.

Fő kutatási kérdésem tehát az, hogy lehet-e az SVM gépi tanulási eszköz előrejelzésére hagyatkozva a vállalati felvásárlások maximum 3 éves határidős piacán a piaci kockázattal korrigált hozamnál szignifikánsan többet elérni, és ha igen, akkor mi a kereskedés algoritmusa. Ez egyben ekvivalens azzal a kérdéssel, hogy hatékony-e az egyesült államokbeli felvásárlási, M&A határidős piac.

2. fejezet

Az akvizíció létrejöttét magyarázó tényezők a pénzügyi és közgazdasági szakirodalomban

Jensen & Ruback (1983) cikke szerint a vállalati felvásárlásokat az egyes vállalatok menedzsmentjének hatékonysága határozza meg: az erőforrásokkal pazarlóan bánó menedzsmenttel rendelkező vállalatok felvásárlásával, és így a hatékonytalan menedzsment leváltásával a felvásárló vállalat hatékony menedzsmentje jelentős többletértéket tud teremteni. Így természetesen a felvásárolt vállalatok részvényesei nyernek a tranzakción, ami felveti a kérdést, hogy miért nem maguk váltották le a hatékonytalan menedzsmentet. A válasz nyitja a megbízó-ügynök probléma, ugyanis a részvényesek, mint megbízók nem feltétlenül rendelkeznek elegendő információval a menedzsment, mint ügynökök tevékenységéről. Így a vezetők nem rendelkeznek elegendő ösztönzéssel sem, hogy hatékonyan irányítsák a vállalatot. A felvásárló vállalatok tehát mintegy kiváltják a részvényesek részéről való monitoringot a felvásárolt vállalat menedzsmentje felett, így hozzájárulnak a piaci hatékonyság növeléséhez is. Jensen és Ruback cikkének nyomán a menedzsment hatékonyságát az eszközarányos árbevétel, az adózott eredmény/árbevétel és EBIT[1]/árbevétel, illetve a részvényegységre jutó cash flow segítségével operacionalizáltam.

Lewellen (1971) szerint a vállalati felvásárlások tradicionálisan elfogadott ösztönzői mellett (szinergiák a két vállalat együttes működéséből) egy felvásárlásnak akár tisztán pénzügyi ösztönzői is lehetnek: amennyiben a felvásárlandó vállalat nem tudja elérni az adósságának megfelelő optimális tőkeszerkezet szintjét, akkor a felvásárló vállalat értéket teremthet annak akvirálásával, és a tőkeszerkezet optimalizációjával. Lewellen munkája nyomán vettem be a modellbe az összes

[1] Earnings Before Interest – számviteli mutató

adósság/saját tőke változót.

Palepu (1986) szerint az úgynevezett növekedés-erőforrás aszimmetria is indokolhatja egy vállalati felvásárlás létrejöttét: a hipotézis szerint az alacsony növekedési potenciállal, de bőséges erőforrásokkal, vagy a jelentős növekedési potenciállal, de szűkös erőforrásokkal rendelkező vállalatok kerülnek felvásárlásra. A potenciálisan akvirálandó vállalat tehát nem rendelkezik a két tényező egyikével, amit a felvásárló vállalat segítségével tud realizálni, ami az értékteremtés forrása. A rendelkezésre álló adatok miatt csak a bőséges erőforrás indikátorát sikerült operacionalizálnom a készpénz/összes eszköz változója személyében.

Peat & Stevenson (2008) egy további lehetséges magyarázó változót vázol fel: a magas könyv szerinti érték/piaci érték aránnyal rendelkező vállalatok nagyobb valószínűséggel kerülnek felvásárlásra. Az előbbi összefüggés arra a tényre épül, hogy a magas könyv szerinti értékkel rendelkező vállalatok feltehetően nagyobb mennyiségű pénzzé tehető eszközzel rendelkeznek, ami korlátozza a felvásárló vállalat számára a potenciális maximális veszteséget.

3. fejezet

Statisztikai módszerek az M&A előrejelzésére a szakirodalomban

3.1. Diszkriminancia-analízisre építő modellek

Stevens (1973) az elsők között alkalmazta a lineáris diszkriminancia elemzés módszerét egy véletlenszerűen kiválasztott vállalat kategorizálására aszerint, hogy vajon várhatóan felvásárlás tárgyává válik-e. A modellben szereplő változókat Stevens a következő öt csoportba sorolta: profitabilitás, likviditás, tőkeáttétel, aktivitás és egyéb változók. Az 3.1. táblázat tartalmazza az öt faktor operacionalizálására használt pénzügyi mutatókat. Természetesen az azonos faktorhoz tartozó változók erősen összefüggenek, így, hogy a multikollinearitásból következő anomáliákat (például a becsült együtthatók nagy érzékenysége a mintára) elkerülje, Stevens (1973)a főkomponens-elemzés módszerét alkalmazta hat egymástól független faktor létrehozására. A létrejött hat faktor közül az első négy megegyezett az 3.1. táblázatban szereplő első négy faktorral, az egyéb kategória pedig további két faktort szolgáltatott az osztalékfizetési ráta és P/E[1] mutatóinak személyében. Mivel a relatív magas loadinggal rendelkező változók adják az egyes főkomponensek varianciájának lényeges részét, így a szerző minden faktort az adott faktorhoz tartozó, legmagasabb loadinggal rendelkező változóval helyettesített. Így vált a profitabilitás, likviditás, tőkeáttétel és aktivitás proxy változójává rendre az $\frac{EBIT}{Árbevétel}$, $\frac{Nettó\,forgótőke}{Összes\,eszköz}$, $\frac{Hosszú\,lejáratú\,kötelezettségek}{Összes\,eszköz}$ és az $\frac{Árbevétel}{Összes\,eszköz}$.

A legmagasabb magyarázóerővel bíró változó a tőkeáttétel lett, majd ezt követte a profitabilitás, aktivitás és végül a likviditás. Az osztalékfizetési ráta és P/E mutatók nem járultak hozzá a diszkriminancia függvény (csoportok közötti eltérés négyzetösszeg/csoporton belüli eltérés négy-

[1]Price/Earnings, vagyis részvényegységre jutó nyereség

3.1. táblázat. Stevens (1973)-as cikkében szereplő magyarázó változók

Faktor	Változók
Likviditás	Nettó forgótőke/Összes eszköz
	Nettó forgótőke/Árbevétel
Profitabilitás	EBIT/Összes eszköz
	Bruttó profit/Árbevétel
	EBIT/Árbevétel
	Adózott eredmény/Árbevétel
	Adózott eredmény/Saját tőke
	Adózozott eredmény/Összes eszköz
Tőkeáttétel	Hosszú lejáratú adósság/Piaci kapitalizáció
	Hosszú lejáratú kötelezettségek/Piaci kapitalizáció
	Hosszú lejáratú adósság/Saját tőke
	Hosszú lejáratú adósság/Összes eszköz
	Összes kötelezettség/Összes eszköz
Aktivitás	Árbevétel/Összes eszköz
	Eladott áruk bekerlési értéke/Készletek
	Árbevétel/Forgóeszközök-Készletek
Egyéb	Kamatok/Készpénz+Értékpapír
	Fizetett oszatllék/Adózott eredmény
	Részvényárfolyam/Részvényegységre jutó nyereség

Forrás: Stevens (1973), 3.1 táblázat, 9. oldal.

zetösszeg) javításához, így ezek nem kerültek be a modellbe. Az optimális modell pontossága (a helyesen kategorizált megfigyelések száma osztva az összes megfigyelések számával) 70%-os lett, aminek az 50%-os, véletlennek megfelelő arányhoz képesti eltérését Stevens t-próbával tesztelte: az eredmények szerint elvethető a pontosság 50%-kal való egyezőségét állító nullhipotézis 0.0005 szignifikancia szint mellett. A modell a 70%-os pontosságot ugyan csak a tanulómintán belül (ami 80 megfigyelést tartalmazott) érte el, ugyanakkor a 80-as mintát egy 40 elemből álló tanulómintára és egy 40 elemből álló tesztelő mintára osztva szintén jelentős, 67.5%-os pontosság adódott. A véletlenhez képest szignifikánsan jobb klasszifikáció felveti a kérdést, hogy lehet-e profitábilis, a piaci hozamhoz képest magasabb kockázattal korrigált hozamot elérő stratégiát konstruálni. Ha a válasz pozitív lenne, az ellentmondana a hatékony piacok hipotézisének. A többváltozós diszkriminancia-analízist Stevens után több szerző is alkalmazta, és a modellek gyakran valóban nem rendelkeztek a véletlennél jobb pontossággal.

Belkaoui (1978) Stevensszel szemben nem egyesült államokbeli, hanem kanadai vállalatok felvásárlásának valószínűségének becslésére tett kísérletet, mintája 25-25 vállalatot tartalmazott. A 3.2 táblázat mutatja, hogy bár változói hasonlítanak Stevens változóihoz, a csoportosításuk mégis eltérő. A „Nem likvid eszközcsoport” első négy változója a profitabilitás faktorához tartozik, ugyanezen csoportnak az ötödik változója, valamint a harmadik eszközcsoport változói a tőkeáttétel faktorhoz tartoznak, míg a második és negyedik változócsoport rendre a likviditás és az aktivitás faktorának feleltethető meg. A szerző a felvásárlás évéhez képest 5 évre visszamenőleg gyűjtötte össze a felvásárolt vállalatok pénzügyi adatait, így meg tudta vizsgálni, hogy milyen időtávra lehet hatékonyan előrejelezni egy vállalati felvásárlás bekövetkezését. Eredménye szerint a lineáris diszkriminancia modell globális tesztjéhez kiszámolt F-statisztika 5%-os szignifikancia szint mellett szignifikanciát jelez minden évre vonatkozóan 4 éven belül, ugyanakor a felvásárláshoz képest 5 évvel korábban megszűnik szignifikáns lenni. A modell pontossága a felvásárláshoz képest öt évvel korábbi időpontot tekintve 72% lett (ez a legalacsonyabb pontossághoz tartozó év), a felvásárláshoz képest három évvel korábban pedig 84% (ami a legmagasabb pontossághoz tartozó év). Fontos ugyanakkor megjegyezni, hogy mind Stevens (1973), mind Belkaoui (1978) cikke alacsony mintaelemszámot vett figyelembe (rendre 80 és 50 elemű minta), így elképzelhető, hogy az eredmények csak a véletlennek köszönhetők. TDK dolgozatomban én is alkalmazom Belkaoui módszerét, és a dolgozat egyik célja annak vizsgálata, hogy 3 éven belül lehet-e ésszerűen előrejelzni egy vállalati felvásárlás létrejöttét.

Fogelberg *et al.* (1978) felhívja a figyelmet egy, a megfigyelésekkel kapcsolatos, alapvető problémára: a mintába tartozó vállalatok, melyek a kutatás időpontjáig még nem váltak felvásárlás tárgyává, elképzelhető, hogy felvásárlás szempontjából igenis kívánatos célpontok lehetnek, ugyanakkor erről a tulajdonságról nem rendelkezünk semmiféle információval. A problémát a szerzőkhöz hasonlóan igyekszünk kezelni, tehát hagyunk egy hosszabb időt eltelni a mintavételezés időpontja és a jelen között, így a felvásárlás tárgyává nem váló vállalatokról nagyobb biztossággal tudjuk állítani, hogy a potenciális M&A célpont indikátor változója mentén 0 értéket érdemelnek. A szerzők új-zélandi vállalatok 86 elemű mintáját alkalmazták a korábbi cikkekhez hasonló többváltozós diszkriminancia modell illesztésére, a változókat pedig Stevens (1973) cikkéhez hasonlóan a faktorelemzés eszközével hozták létre. A modell hasonló változókat tartalmaz, mint az előzőleg ismertetett cikkek (profitabilitás, likviditás, külső várakozások és részvényegységre jutó nyereség változékonysága), ugyanakkor a modell pontossága alacsonyabb, mindössze 60%. A szerzők ugyanakkor hangsúlyozzák, hogy a diszkriminancia modell összes együtthatója intuitív lett: magasabb profitabilitású, likviditású és a külső várakozások mentén magas értéket felvevő (lényegében magas P/E aránnyal rendelkező) vállalatok kisebb eséllyel válnak vállalati felvásárlás tárgyává, illetve a részvénységre jutó nyereség magas változékonysága a felvásárlás esélyének csökkenését vonja maga után. Az első három változóval kapcsolatos eredmény szinkronban van

3.2. táblázat. Belkaoui (1978)-as cikkében szereplő magyarázó változók

Faktor	Változók
Nem likvid eszközcsoport	Cash flow/Nettó vagyon
	Cash flow/Összes eszköz
	Adózott eredmény/Nettó vagyon
	Adózott eredmény/Összes eszköz
	Hosszú lejáratú hitelek+Elsőbbségi részvények/Összes eszköz
Likvid eszközök aránya összes eszközön belül	Forgó eszközök/Összes eszköz
	Készpénz/Összes eszköz
	Nettó forgőtőke/Összes eszköz
	Készpénz+Értékpapírok+Követelések/Összes eszköz
Likvid eszközök és rövid lejáratú kötelezettségek aránya	Rövid lejáratú eszközök/Rövid lejáratú kötelezettségek
	Készpénz+Értékpapírok+Követelések/Rövid lejáratú kötelezettségek
	Készpénz/Rövid lejaratú kötelezettségek
Likvid eszközök forgási sebessége	Forgóeszközök/Árbevétel
	Készpénz+Értékpapírok+Követelések/Árbevétel
	Nettó forgótőke/Árbevétel
	Készpénz/Árbevétel

Belkaoui (1978), 95. oldal, 1. táblázat

a pénzügyi szakirodalmi részben ismertetett, Jensen & Ruback (1983) nevéhez kötődő elmélettel, vagyis a hatékonyabb, jobb vezetéssel rendelkező cégeket kisebb eséllyel vásárolják fel, hiszen a felvásárlással együtt járó menedzsmentcserével kevesebb érték teremthető.

Rege (1984) hasonló következtetésre jut, mint Fogelberg *et al.* (1978): diszkriminancia modell globális tesztjéhez kiszámolt F-statisztika nem jelzett szignifikanciát. Rege (1984) modellje a korábban ismertetett cikkekhez hasonló pénzügyi mutatókat tartalmaz (likviditás, tőkeáttétel, osztalékkifizetés, aktivitás és profitabilitás), a minta pedig 110 kanadai vállalatot foglal magába. A szerző hangsúlyozza, hogy ha sikerülne pusztán pénzügyi arányszámokra építve egy szignifikáns diszkriminancia modellt építeni, az ellentmondana a hatékony piacok hipotézisének (azon belül is a közepesen erős változatának, amennyiben a pénzügyi arányszámokat fundamentális információnak tekintjük). Ennek fényében tehát a modell inszignifikáns magyarázó ereje nem meglepő. Egy másik lehetséges gondolat, ami magyarázza a pénzügyi mutatók inszignifikanciáját, hogy a vállalati felvásárlás nem azok aktuális értékétől, hanem jövőbeli várható alakulásától függ. Ugyanakkor, ahogy a szerző is kiemeli, a jövőbeli várakozásokra vonatkozó adatok nagyon nehezen érhetők el, és minőségük mindig kérdéses.

3.2. Egy probit modell

Harris *et al.* (1982) az imént bemutatott négy tanulmánnyal ellentétben nem a diszkriminancia elemzés, hanem a probit regresszió módszerét alkalmazza. A probit modellnek kétféle változatát is illesztik: a fix együtthatós probit modell konstans lineáris kombinációs együtthatókkal rendelkezik, a véletlen együtthatós probit modell esetén pedig maguk az együtthatók is valószínűségi változók. A véletlen együtthatós probit modell pénzügyileg úgy is értelmezhető, hogy a felvásárló vállalattól függ, hogy a felvásárolt vállalat egy adott pénzügyi jellemzője milyen erős, vagy akár milyen előjelű hatást gyakorol a felvásárlás valószínűségére. Például egy magas tőkeáttétellel rendelkező vállalat számára az alacsony tőkeátétellel rendelkező felvásárlandó vállalat értékes célpont lehet az eladósodottsági mutatójának csökkentése érdekében, ugyanakkor egy jelenleg alacsony tőkeáttéllel rendelkező, a kamat adópajzsának növelése érdekében adósságának növelésére törekvő vállalat éppen a magas tőkeáttétellel rendelkező vállalat akvizíciójában érdekelt. Így tehát megfordulhat a felvásárlás valószínűsége és a magyarázó változó közötti összefüggés előjele. Harris *et al.* (1982) Stevens (1973) előzőleg bemutatott cikkének kritikájaként jegyzik meg, hogy önkényesen azonos arányban határozza meg a felvásárolt és az akvizíció tárgyává nem váló vállalatok mintabeli halmazát, ugyanakkor a valóságban a felvásárolt vállalatok jóval kisebb arányban fordulnak elő. A szerzők mintájában a felvásárolt vállalatok aránya így jobban tükrözi azok populációbeli arányát (61 felvásárolt vállalat az 1260-ból, tehát 4.8%).

A modellbe bevont változókat a szerzők két csoportra osztják: pénzügyi jellemzőkre és iparági jellemzőkre, szemben a korábban bemutatott cikkekkel, amik csak az előbbi tulajdonságokra helyezték a hangsúlyt. Az iparági jellemzőkön belül a szerzők négy változót vizsgáltak: a négy legnagyobb cég piaci részesedése az adott iparágon belül, reklámköltségek aránya az árbevételen belül, K+F költségek aránya az árbevételen belül és az iparági teljes árbevétel átlagos növekedési rátája 1968 és 1972 között. A fix együtthatós modellben két változó bizonyult mindössze szignifikánsnak mindkét vizsgált periódusban (1974-1975 és 1976-1977): a P/E mutató és a vállalat mérete (ez utóbbit az eszközök állományának logaritmusával operacionalizálták a szerzők), mindkét változó negatív előjellel függ össze a felvásárlás valószínűségével. Az eredmények intuitívnek tekinthetők, hiszen az alacsony növekedési potenciállal (alacsony P/E arány) rendelkező vállalatok feltehetően korrelálnak a menedzsment alacsony hatékonyságával, így Jensen & Ruback (1983) elmélete szerint nagyobb eséllyel válnak felvásárlás tárgyává. A nagyobb vállalatok akvizíciója pedig több tőkét igényel, illetve akár a gazdasági versenyhivatal is megakadályozhatja esetükben az akvizíció létrejöttét, így ezen változó esetén is intuitívnek tűnik a negatív előjel.

A szerzők a korábban bemutatott cikkekkel ellentétben nem alkalmaznak tévesztési (kontingencia) mátrixot, hivatkozva a probit modell outputjának mérési skálájára (a felvásárlás valószínűségére vonatkozik a becslés), bár egy akár némileg önkényesen kijelölt vágási küszöbbel a modell előrejelzése indikátor változóvá alakítható. A tévesztési (kontingencia) mátrix és annak pontosságának kiszámítása helyett a mintán belüli felvásárolt vállalatok arányát (ami 4.8% az 1974-1975 periódusban) összehasonlítják a modell outputjával azokban a hipotetikus esetekben, ha a megfigyelés magyarázó változói éppen a felvásárolt, illetve a fel nem vásárolt vállalatok átlagos jellemzőivel egyeznek meg: így az átlagos felvásárolt vállalat felvásárlásának valószínűsége a modell szerint 6.621%, az átlagos fel nem vásárolt vállalathoz tartozó valószínűség pedig 3.265%[2]. A véletlen együtthatós probit modellen belül a P/E mutató és vállalatméret szintén szignifikánsnak bizonyult a fix együtthatós változatnak megfelelő előjellel, ugyanakkor két további szignifikáns magyarázó változó is bekerült ebbe a modellbe: a felvásárolt vállalat múltban fehalmozott, adóalap csökkentésre felhasználható veszteségének állománya és a korábban már látott aktivitás (árbevétel/összes eszköz). Az előbbi változó meglepő módon negatív előjellel rendelkezett, amit a szerzők azzal magyaráznak, hogy jogilag könnyebb a felvásárló vállalat felhalmozott veszteségét a felvásárolt vállalat adóalapjának csökkentésére fordítani, mint fordítva, bár ebből az érvelésből inkább a változó inszignifikanciája következne. Az aktivitás Jensen & Ruback (1983) elméletének megfelelően negatív koefficienset kapott a modellben, tehát az alacsonyabb hatékonysággal rendelkező vállalatok nagyobb eséllyel válnak felvásárlás tárgyává (amit az elmélet szerint hatékony vállalatok viteleznek ki). Fontos, hogy a véletlen együtthatós modellben az aktivitás együtthatójának szórása a 0-nál szignifikánsan nagyobbnak bizonyult, tehát az aktivitás hatásának erőssége a felvásárlás valószínűségére véletlenszerűen változik, például a felvásárló vállalat személyétől függően. Harris *et al.* (1982) a diszkriminancia elemzés eszköze helyett a probit modellt alkalmazták, ugyanakkor a szakirodalomban az általánosított lineáris modell egy másik változata, a logit modell (logisztikus regresszió) vált elterjedtebbé. A következőkben az ezen módszert alkalmazó munkákat tekintjük végig.

3.3. Logit modellek

Dietrich & Sorensen (1984) a logisztikus regresszió módszerét alkalmazta az M&A sikerességi valószínűségének becslésére. A módszer előnye a korábban bemutatott cikkek által alkalmazott diszkriminancia-analízissel szemben, hogy nem követeli meg előfeltételként a magyarázó változók többdimenziós normális eloszlását (aminek ellenőrzése nem triviális feladat, a koordinátánkénti normalitás szükséges feltételét lehet ellenőrizni). Ugyanakkor a logisztikus regresszió illesztésénél

[2]Tehát modelljük Tjur-féle R^2 mutatója 3.356%, ami meglehetősen alacsony.

fontos feltétel, hogy a megfigyelések a magyarázó változókra feltételesen függetlenek legyenek, hiszen a paraméterek becslése a maximum likelihood elvével történik, és a likelihood függvény a megfigyelések függetlensége mellett írható fel egyszerűen. A függetlenség feltételét azonban a szerzők nem vizsgálják. A cikk különlegessége, hogy egy egyszerű, de logikus elméleti modellt vázol fel, ami alapján megfogalmazza az egyes magyarázó változók és a felvásárlás valószínűsége közötti kapcsolat várható irányát. Nevezetesen egy diszkontált cash flow modellről van szó, az M&A tranzakió létrejöttével együtt járó inkrementális cash flow-k diszkontáltjának összege adja a felvásárló vállalat célfüggvényét, ami ha pozitív, akkor az az M&A tranzakció létrejöttét jelenti.

A szerzők tíz változót vesznek figyelembe, ebből csak kettő olyan van, amit a korábban bemutatott elméleti, illetve statisztikai cikkek valamelyike ne vett volna figyelembe: ez a befektetési hányad (Tőkekiadások/Összes eszköz) és a tőzsdei kereskedés volumene az akvizíció évében. Mindkét változó együthatójának előjele várhatóan pozitív: a magasabb befektetési hányaddal rendelkező, felvásárlás tárgyává váló cégek várhatóan több cash flow-t termelnek, a likvidebb részvénnyel rendelkező vállalatokat pedig kisebb tranzakciós költségek mellett lehet felvásárolni, ami az akvizícióval járó nettó pénzáramokat növeli. A végső modellbe végül öt változó került be: az eszközhatékonyság (Árbevétel/Összes eszköz), vállalatméret (itt Piaci kapitalizáció), tőkeáttétel (Hosszú lejáratú adósság/Összes eszköz), osztalékkifizetési ráta és a kereskedés volumene az akvizíció évében. Az utóbbi változó kivételével mindegyik változó együtthatója szignifikánsan különbözött 0-tól és negatív lett, míg a kereskedés volumenéhez tartozó béta együttható nem szignifikáns, és pozitívan függ össze az akvizíció valószínűségével, a várakozásoknak megfelelően. Az eszközhatékonyság együtthatójának abszolút értéke nagyon látványosan kiugrott a többi változóhoz képest, ami szinkronban van Jensen & Ruback (1983) modelljével: az elmélet szerint az alacsony hatékonyságú vállalatok felvásárlásával a hatékony vállalatok jelentős többletértéket tudnak generálni a hatékonytalan menedzsment leváltásával, ami beleillik a Dietrich & Sorensen (1984) által a cikkben felvázolt diszkontált cash flow modellbe.

A modell pontossága a mintán kívüli teszt szerint 91%-os lett, bár fontos megjegyezni, hogy ez a cikk is viszonylag alacsony mintaelemszámra épít: a modell illesztéséhez használt minta mindössze 24 felvásárolt és 43 fel nem vásárolt vállalatot tartalmaz az 1969-1973-as időszakból, a mintán kívüli tesztet pedig 22 vállalatra alapozzák a szerzők, amiből 6-ot vásároltak fel. Az alacsony mintaelemszám mellett feltűnő, hogy a felvásárolt vállalatok aránya a sokaságon belüli arányhoz képest a mintán belül jóval magasabb.

Palepu (1986) több, a korábban bemutatott cikkekkel is kapcsolatos hibára hívja fel a figyelmet: mivel a mintavételezés úgy történt, hogy egyenlő számban kerültek felvásárolt és fel nem vásárolt vállalatok a mintákba, sérül a megfigyelések függetlenségének feltétele, ami miatt a statisztikai modellünk együtthatóira és magára a felvásárlás valószínűségére adott becslések asszimptotikusan torzítottak és inkonzisztensek lesznek. A felvásárlás valószínűségére vonatkozóan a torzítás azt jelenti, hogy ha a felvásárolt vállalatok mintabeli aránya nagyobb, mint a sokasági aránya, akkor a modellünk által adott becslés a felvásárlás valószínűségére várható értékben nagyobb lesz, mint a tényleges valószínűség. Ez alapján tehát elmondható, hogy a korábban bemutatott tanulmányok a felvásárlás állapotától függő, azonos elemű minták miatt szisztematikusan felülbecsülték egy vállalat felvásárlásának valószínűségét. A paraméterek torzítottságának és inkonzisztenciájának problémája a feltételes vagy súlyozott maximum likelihood módszereinek alkalmazásával oldható fel, míg a felvásárlás valószínűségére vonatkozó torzítás az alkalmazott vágási küszöbnek a sokasági arányhoz való igazításával érhető el. Továbbá ha a teszthalmazunk is egyenlő arányban tartalmaz felvásárolt és fel nem vásárolt vállalatot, akkor a tévesztési mátrix nem reprezentálja megfelelően a valóságos döntési helyzetekben elérhető tévesztési mátrixot. A harmadik kritika a vágási küszöbbel kapcsolatos: annak megválasztásánál ugyanis célszerű figyelembe venni a felvásárolt és fel nem vásárolt vállalatok alapsokaságon belüli arányát, illetve hogy milyen súlyt tulajdonítunk az első és másodfajú hibáknak.

Palepu (1986) nyomán törekedtem arra, hogy a felvásárolt és fel nem vásárolt vállalatok aránya az alapsokaságéhoz hasonló legyen, éppen ezért nem azonos arányban vettem mintát a kétféle vállalatból. A harmadik, Palepu (1986) által hangsúlyozott módszertani probléma a vágási küszöbbel kapcsolatos: egy valós döntési szituációban el kell tudnunk dönteni, hogy a modell által adott becslés a felvásárlás valószínűségére elegendő-e ahhoz, hogy az adott vállalatot felvásárlásra alkalmasnak tekintsük. Ki kell tehát választanunk egy vágási küszöböt, ami feletti felvásárlási valószínűséggel rendelkező vállalatokat felvásárlandónak kategorizáljuk. Palepu megjegyzi, hogy a cikkéhez képest korábbi szakirodalmak nem vizsgálták a vágási küszöb optimális megválasztását, helyette általában az esetleges 0.5-ös értéket választották. A vágási küszöb megválasztásakor ugyanakkor figyelembe kell venni a felvásárolt és fel nem vásárolt vállalatok arányát, illetve az első és másodfajú hiba költségét.

A szerző a modellt 163 felvásárolt és 256 fel nem vásárolt vállalat mintája alapján illeszti, a vállalatok az 1971-1979-es időszakból származnak az Egyesült Államokból. Látható, hogy a célváltozó szerint különböző vállalatok aránya nem egyezik az alapsokasági aránnyal, a felvásárolt vállalatok felülreprezentáltak a fel nem vásárolt vállalatokhoz képest. Ezért a szerző korrigálja a logisztikus regresszió tengelymetszet paraméterét a felvásárlás valószínűségének torzítatlan becslése érdekében. A logisztikus regressziós modell tesztelésére választott teszthalmaz pedig az alapsokasági arány figyelembe vételével kerül kiválasztásra, mindössze 30 felvásárolt vállatot tartalmaz az 1 117-es számosságú tesztminta. Palepu modellje hasonló magyarázó változók bevonására tett kísérletet, mint a korábban bemutatott szakirodalmak, végül három változó bizonyult szignifikánsnak: a menedzsment hatékonyságát kifejező éves átlagos többlethozam (amit a szerző egy piaci indexmodellel operacionalizál), a vállalatméret és az erőforrás-növekedés aszimmetriát kifejező dummy változó. Az erőforrás-növekedés aszimmetria elméletét az első fejezetben már megvizsgáltuk, lényege, hogy az alacsony növekedési potenciállal és bőséges erőforrásokkal vagy a magas növekedési potenciállal és szűkös erőforrásokkal rendelkező vállalatok nagyobb eséllyel válnak akvizíció tárgyává.

A modell használatához a szerző először is meghatározza az optimális vágási küszöböt: ehhez felvázolja felvásárlás tárgyává való válás sűrűségfüggvényét (pontosabban annak diszkrét közelítését a modell illesztésére használt minta alapján) mind a felvásárolt, mind a fel nem vásárolt vállalatokon belül, és a két sűrűségfüggvény metszéspontjaként határozza meg az optimális vágási küszöböt. E választás kiindulópontja, hogy az első és másodfajú hibát egyenértékűnek tekintjük, így ha a vágási küszöb a két sűrűségfüggvény metszéspontja alatt lenne, akkor azt magasabbnak választva nagyobb mértékben csökkentjük a tévesen kategorizált vállalatok számát a felvásárolt vállalatokon belül, mint amennyivel nő a tévesen besorolt vállalatok száma a fel nem vásárolt vállalatok halmazában. Hasonló érveléssel lehet látni, hogy az optimális vágási küszöb a két sűrűségfüggvény metszéspontja fölött sem lehet, így éppen a metszéspontot kell választanunk. Az 1 117 elemből álló teszthalmazon a modell pontossága 45.6%-os lett, és a ténylegesen felvásárolt 30 vállalaton belül a modell 24-et talált meg, vagyis itt 80%-os pontosság mutatható ki. Ennek ugyanakkor az az ára, hogy a modell a fel nem vásárolt vállalatok 55%-át tévesen felvásároltnak klasszifikálja.

Felmerül a kérdés, hogy vajon az eredmények ellentmondanak-e a hatékony piacok hipotézisének, lehet-e ilyen valószínűségek mellett profitábilis stratégiát konstruálni? A válasz természetesen azon múlik, hogy milyen várható többlethozamot jelent, ha egy vállalatot felvásárolnak. A 24 helyesen klasszifikált felvásárolt vállalat éves átlagos többlethozama (a hasonló bétával rendelkező részvényekhez képest) a felvásárlásig 20.98% volt (Palepu (1986) alapján), míg a tévesen felvásároltnak kategorizált, de fel nem vásárolt vállalat többlethozama csak -1.62%. Tekintve, hogy a 24 felvásárolt vállalat beazonosításához 625 részvényben kell long pozíciót felvennünk, nehezen hihető, hogy a modell arbitrázsra adna lehetőséget, amit Palepu is hangsúlyoz. Fontos megjegyezni, hogy a vágási küszöb megállapítása az első és másodfajú hiba ekvivalensnek való feltételezése mellett került megállapításra, ugyanakkor a felvásárolt és a fel nem vásárolt cégekhez tartozó többlethozamok éppen arra engednek következtetni, hogy egy felvásárolt vállalat beazonosításának elszalasztása súlyosabb költséggel jár, mint egy fel nem vásárolt vállalat félrekategorizálása, illetve a vágási küszöb megállapítása a sokaságon belüli arányok figyelembe vétele nélkül történt.

4. fejezet

A Support Vector Machine

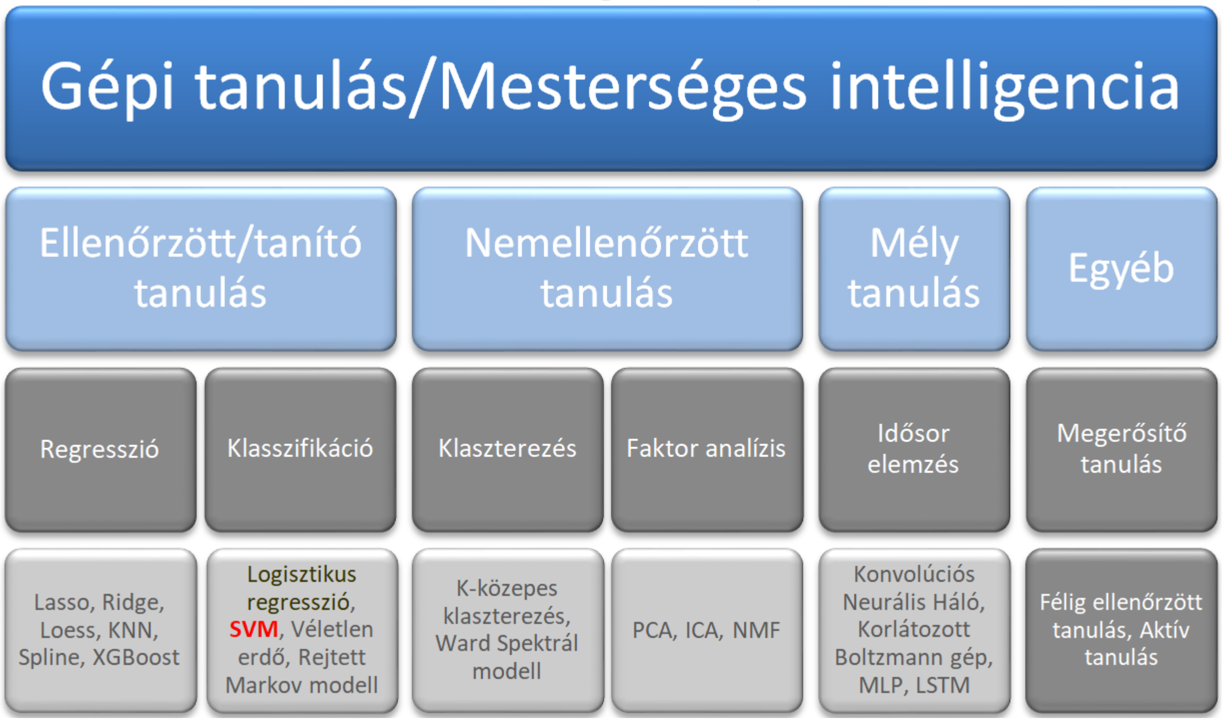
4.1. Általános tudnivalók

A Support Vector Machine (továbbiakban SVM)[1] egy gépi tanulási eszköz, mely feladata klasszifikálni bináris eredményváltozójú megfigyeléseket a lehető leghatékonyabban. A fejezet ezen első mondata már azonnal magyarázatra szorul. Kezdjük a végéről: a hatékonyság jellemzésére rengeteg mérőszám létezik. Ott van például a mindenki által ismert R^2 mutató, vagy híres információs kritériumok, mint például az Akaike, vagy a BIC. A mérőszámok haszálata azonban feltételekhez kötött, és nem lehet bármilyen statisztikai vizsgálathoz ugyanazt rendelni. Példának okáért, a logisztikus regresszió illeszkedését nem lehet a fentebb említett R^2 számmal jellemezni, ugyanis annak a kiszámításához léteznie kell belső, illetve külső varianciafelbontásnak, de a bináris klasszifikációknál ez nem áll fent. Sajnos ugyanez igaz az SVM-re is. A formalizmus részletezésénél így külön kitérek erre a pontra is. A gépi tanulást, angolul Machine Learninget, sokan tévesen már a mesterséges intelligencia egyik megtestesüléseként fogják fel. Azonban nem szabad elfelejtkeznünk egy dologról: a mesterséges intelligencia is csak emberi találmány, matematikával és informatikával leírható folyamatok összessége. Statisztikai szemszögből vizsgálva azt mondhatjuk, hogy van az úgynevezett hagyományos iskola, ami a különböző típusú regressziókat, és azok modellépítési eljárásait illetve feltételrendszereit vizsgálja, illetve van a modern iskola, aminek az alapja ugyanez, csak inkább algoritmikus módon, informatikai oldalról közelít. Ezen pályamunka célja nem az, hogy a gépi tanulás minden részletét pontosan leírja, csak az, hogy elhelyezze egy keretbe az SVM-et. Így a teljesség igénye nélkül a Machine Learning eszközök osztályzása:[2]

[1] A módszer magyar neve Tartóvektor Gép, azonban ez még a magyar szakirodalomban sem használt kifejezés.

[2] Hosszabb cikk a témáról: https://news.efinancialcareers.com/uk-en/285249/machine-learning-and-big-data-j-p-morgan

4.1.1. ábra. Gépi tanulás fajtái



Forrás: JP Morgan Macro QDS

4.2. Matematikai formalizmus

A Support Vector Machine elméletét Vapnik (2013) dolgozta ki, a levezetéseket a cikke alapján mutatom be.

4.2.1. Lineárisan szeparálható minta

1. Definíció. Azt mondjuk, hogy egy n magyarázóváltozóval rendelkező minta lineárisan szeparálható, ha létezik a magyarávoltozók által kifeszített térben n-1 dimenziós maximális altér (hipersík), mely úgy particionálja a teret, hogy az egyik particióban csak az egyik, a másik particióban pedig csak a másik kategóriájú megfigyelésvektor tartózkodik.

A definíció közvetlen következményeként látható, hogy lineárisan szeparálható mintákat tekintve a szeparáló hipersíkok halmaza nem üres. Így legalább egy ilyen létezik, ám általában e halmaz számossága kontinuum. Egy dolgot azonban fontos kiemelni: az ilyen hipersíkok meredekségei[3] nem mehetnek egy adott véges (mintától függő) érték fölé, illetve egy adott véges érték alá[4]. Így a meredekségek *korlátos és zárt intervallumot alkotnak a valós számok terében*. Ha nem alkotnának korlátos halmazt, akkor triviálisan nem lehetne szeparáló hipersík, aminek a lehetséges meredekségeit tekintjük (ekkor ugyanis úgymond meg lehetne forgatni a hipersíkot úgy, hogy továbbra is szeparáló maradjon, de ekkor lehetne mutatni olyan állást, hogy a partició nem történik meg. Ez ellentmondás.)

2. Definíció. Azt mondjuk, hogy egy szeparáló hipersík optimális, ha a megfigyelésvektorok hipersíktól vett minimális távolsága maximális. Ezen távolság kétszeresét a szakirodalom *marginnak* hívja. Így optimális egy hipersík, ha a hozzátartozó margin maximális.

Érzékeltetés végett, két dimenziós térben valahogy a következően néz ki a történet. Tegyük fel, hogy lepkéket akarunk osztályozni, mégpedig úgy, hogy elmegyünk az erdőbe, és fogunk pár lepkét. Tegyük fel hogy két kategóriája van az erdőben megfigyelhető lepkéknek, mondjuk csodalepke és csúnyalepke. Az állatoknak mind megmérjük a testhosszát és a fesztávolságát, majd feljegyezzük ezt a következő módon: $\left[fesztávolság \quad testhossz \quad kategória\right] = \left[2mm \quad 6mm \quad csodalepke\right]$. Egyszer csak egy számunkra ismeretlen lepkét fogunk. Megmérve az adatait akarjuk bekategorizálni. Ebben segít nekünk az SVM. Lineáris szeparábilitás esetében ez a 4.2.1 ábrának megfelelően néz ki.

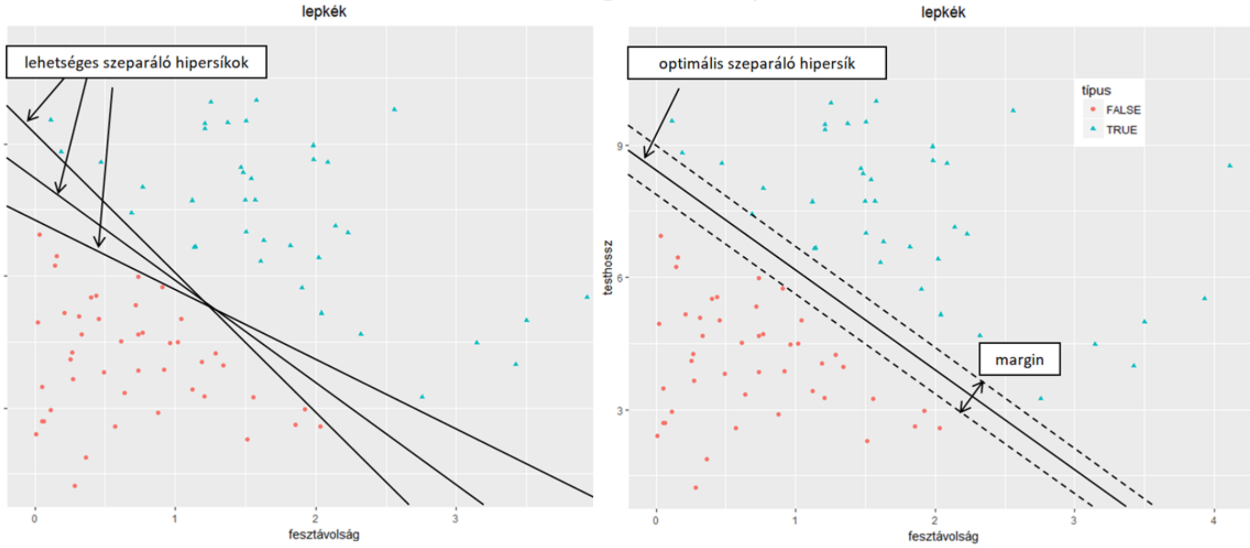
Amennyiben az optimális egyenestől jobbra esik a lepkénk a tulajdonságai alapján, akkor TRUE azaz csodalepke kategóriát kap, amennyiben balra, akkor pedig FALSE, azaz csúnyalepke kategóriát. Félreklasszifikálás esélyének minimalizálása végett szükséges megkeresnünk az optimális hipersíkot, ami a képen látható. A példa bugyuta, de nagyon sok mindenre rá lehet húzni: banki csőd becslés (egy ismeretlen bank csődbe menetelét előrejelezni ismert minta alapján), egészségügyi becslések (például ellenszérum hatásosságának becslése ismerve a páciens attribútumait), és még sok más.

Térjünk vissza a matematikára. A lehetséges hipersíkok meredekségei a fentiek alapján kompakt halmazt alkotnak (teljes téren a kompaktság ekvivalens a korlátosság és a zártság egyszerre teljesülésével, és a valós tér ilyen). Vegyük észre: egy hipersík adott, ha a meredeksége, és egyik tengelymetszete adott. Intuitíve egyértelmű, hogy egy adott tengelyhez tartozó metszetek lineárisan szeparálható minta esetén ismét korlátos és zárt halmazt alkotnak. Így a meredekségek és tengelymetszetek Descartes-szorzatán értelmezett margin (távolság) függvény a folytonossága miatt Wei-

[3]Meredekség alatt a bázisvektorokkal bezárt szög tangensét értem. Két dimenzióban elképzelve ez megegyezik a meredekségről alkotott képünkkel.

[4]Itt egy kis sarkítás történt. Van mikor az említett meredekség $+\infty$. Ilyenkor azonban létezik bázis, amire tekintve ez a meredekség egy véges szám. Áttérve erre a bázisra megoldottuk a problémát.

4.2.1. ábra. Lineáris szeparálhatóság 2 dimenzióban



Saját ábra

ersraß tétele alapján *felveszi a maximumát*. Összeségében tehát elmondható, hogy egy lineárisan szeparálható SVM feladatnak mindig van optimális megoldása. További feladatunk megkeresni ezt az optimumot.

Nevezzünk mintának egy $\{x_i, d_i\}_{i=1}^{p}$ halmazt, ahol x_i egy adott $i = 1, 2, \ldots, p$-hez tartozó magyarázóváltozó vektor, d_i pedig az ugyanehhez tartozó kategórikus változó. A későbbi egyszerűség kedvéért tegyük fel, hogy $d_i \in \{1, -1\}$. Jelölje ω a keresett hipersík normáltvektorát, és b a hipersík egyenletében lévő konstans tagot (tengelymetszetet) [5]. Vezessük be a $g : X \to \mathbb{R}$ függvényt, ahol $\forall x_i \in X$, $g(x_i) = \omega^T \cdot x_i + b$. (A továbbiakban a $\cdot$ jel a vektorszorzásra utal, míg a skalárral való szorzást úgy jelölöm, hogy egyszerűen elhagyom a szorzásjelet). Matematikai formalizmussal élve azt mondhatjuk, hogy a minta lineárisan szeparálható, ha

$$g(x_i) = \omega^T \cdot x_i + b \begin{cases} > 0 & ha\ d_i = 1 \\ < 0 & ha\ d_i = -1 \end{cases} \tag{4.2.1}$$

Ezt másképpen a döntési szabálynak nevezzük. Tehát ha egy eddig ismeretlen elem kategóriája a kérdés, akkor megnézve a g függvény általi képet dönthetünk: ha pozitív, akkor egyes kategóriájú, ha negatív, akkor pedig mínusz egyes.

[5] Kis magyarázat: Képzeljük el azt az egyenest a két dimenziós síkon, amelyik a -1-ben metszi az y tengelyt, és $\frac{3}{2}$ a meredeksége. Ennek az egyenesnek az egyenlete $2y = 3x - 2$. Ezt átrendezve kapjuk a $0 = 3x - 2y - 2 = \begin{bmatrix} -3 & 2 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} - 2$ úgynevezett normálvektoros egyenletet. Ennek az egyenesnek a normálvektora a $\omega = \begin{bmatrix} -3 \\ 2 \end{bmatrix}$ vektor, tengelymetszete pedig a $b = 2$ szám.

Mivel tetszőleges két valós szám között található valós szám (a valós számok sűrűk a számegyenesen), ezért létezik olyan ε pozitív valós szám, hogy

$$\begin{cases} g(x_i) \geq \varepsilon > 0 & ha\ d_i = 1 \\ g(x_i) \leq -\varepsilon < 0 & ha\ d_i = -1 \end{cases}$$

Ezzel az ε számmal leosztva az egyenlőtlenséget (összehúzzuk a teret, avagy átkoordinatizáljuk), illetve felhasználva, hogyha ω egy normálvektor, akkor annak tetsőleges skalárszorosa is az, kapjuk, hogy

$$\begin{cases} \overline{\omega^T} \cdot x_i + \bar{b} \geq 1 & ha\ d_i = 1 \\ \overline{\omega^T} \cdot x_i + \bar{b} \leq -1 & ha\ d_i = -1 \end{cases}$$

Összefoglalásképpen,egy lineárisan szeparáló hipersík lehetséges megoldása az SVM-nek, amennyiben normálvektorai közül az egyikre igaz, hogy

$$d_i(\overline{\omega^T} \cdot x_i + \bar{b}) \geq 1,\ x_i \in X,\ i = 1, 2, \ldots, p \tag{4.2.2}$$

Mostmár csak meg kell fogalmaznunk egy célfüggvényt, amit ezen korláton optimalizálunk, és készen is vagyunk.

Az első definició (1) alapján a hipersík egy altér. Így a projekciós tétel (Puskás & Dancs (2002)) alapján a magyarázóváltozók által kifeszített X vektortér előáll mint a hipersík és annak az ortogonális kiegészítőjének direkt összege ($X = H \bigoplus H^{\perp}$). Így

$$x_i = x_{k_i} + \frac{\omega}{\|\omega\|} r_i,\ x_{k_i} \in H,\ \omega \in H^{\perp},\ r_i \in \mathbb{R},\ x_i \in X,\ i = 1, 2, \ldots, p \tag{4.2.3}$$

Itt felhasználtuk azt a tényt, hogy H egy hipersík, így maximális altér, így a direkt kiegészítője egy dimenziós. Ez nem jelent mást, minthogy a normálvektora generálja: $H^{\perp} = lin(\omega)$. Tehát minden H-ra merőleges vektor $r\omega,\ r \in \mathbb{R}$ alakú.

Használjuk fel az előző összefüggést döntési szabályunk (4.2.1) ellenőrzésekor!

$$g(x_i) = \omega^T \cdot x_i + b = \omega^T \cdot (x_{k_i} + \frac{\omega}{\|\omega\|} r_i) + b = \omega^T \cdot x_{k_i} + b + \omega^T \cdot \omega \frac{1}{\|\omega\|} \cdot r_i = 0 + r_i \frac{\|\omega\|^2}{\|\omega\|}$$

$$g(x_i) = \|\omega\| r_i$$

Ebből egyből kapjuk, hogy $r_i = \frac{g(x_i)}{\|w\|}$. Tehát nem csináltunk mást, mint a hipersík körüli pontokat felbontottuk két összetevőre, majd beláttuk, hogy a döntési szabályunkat (4.2.1) alkalmazva rá a g függvény általi kép csak a hipersíkra merőleges összetevőtől függ. Továbbá, egy pont hipersíktól való távolsága megegyezik a pont és a merőleges vetítésének különbségének hosszával (ami a ω hosszának számszorosa), ami viszont pont a hozzá tartozó r valós szám abszolút értéke! A projekciós egyenletünkben (4.2.3) ezért osztottuk le a normálvektort a hosszával, mert így az egységnormálvektor számszorosait tekintettük.

A margin optimális nagyságának meghatározásakor (2. definició) a hipersíkhoz legközelebb lévő pont távolsága a maximalizálandó. Legközelebb van egy pont, ha a hozzá tartozó $|r|$ minimális, tehát a lehetséges megoldásokról szóló egyenlet alapján (4.2.2) akkor, ha $r = \frac{g(x)}{\|\omega\|} = \frac{1}{\|\omega\|}$, vagy $r = \frac{g(x)}{\|\omega\|} = \frac{-1}{\|\omega\|}$. Ilyen pontokat tekintve a margin nem más, mint $m = \frac{1}{\|\omega\|} - \frac{-1}{\|\omega\|} = \frac{2}{\|\omega\|}$. Fontos még megjegyezni, hogy $g(x) = \pm 1,\ ha\ \omega^T \cdot x + b = \pm 1$.

3. Definíció. Azokat a pontokat (vektorokat), amelyekre $\omega^T \cdot x + b = \pm 1$, tartóvektoroknak, másnéven *support vektoroknak* nevezzük.

Emlékezzünk vissza: azt a szeparáló hipersíkot keressük, amelyhez tartozó margin maximális, vagyis amelyhez képest a pontok minimális távolsága is maximális. Így az optimalizálási feladatban $\frac{2}{\|\omega\|}$-t maximalizáljuk, de helyette egy ekvivalens feladatot oldunk meg:

$$\begin{cases} min: & \|\omega\| \\ & \\ & d_i(\omega^T \cdot x_i + b) \geq 1 \\ feltéve: & d_i \in \{-1, 1\} \\ & i = 1, 2, \dots, p \end{cases} \tag{4.2.4}$$

A két célfüggvény ekvivalens egymással abban az értelemben, hogy a kettőnek ugyanazon ω-nél lesz optimuma (hiszen egymás reciprokainak konstansszorosai). Egyetlen szingularitási pont a $\omega = 0$ lehetne, ámde $H^\perp = lin(\omega)$, tehát ω bázisvektor. Bázis pedig nem lehet a nullavektor, így ez nem jelent problémát. A minimalizálandó függvény konvex, ugyanis teljesíti a konvex függvények definícióját, miszerint a függvény tetszőleges két pontja felett halad el az őket összekötő szelő: $\|\lambda u + (1-\lambda)\upsilon\| \leq \|\lambda u\| + \|(1-\lambda)\upsilon\| = \lambda\|u\| + (1-\lambda)\|\upsilon\|, \forall u, \upsilon \in X, , \lambda \in [0,1]$.[6] Ez egyébként nem meglepő, tudva, hogy a normafüggvény az általános távolságfüggvénynek felel meg, ami viszont könnyen elképzelhető két dimenzióban. Ekkor bármely háromszög bármely oldala rövidebb mint a másik kettő hosszának összege, ez pedig a távolságfüggvény két dimenziós konvexitásának igazolása. Továbbá, amennyiben az Euklideszi normát használjuk, akkor a célfüggvény

[6]Az első egyenlőtlenség az úgynevezett háromszög-egyenlőtlenség, amit minden normafüggvény teljesít definíció szerint. A második reláció pedig azért teljesül, mert a normából a pozitív skalár ilyen módon kiemelhető.

nemcsak hogy konvex, de kvadratikus is[7]. Ugyanolyan fontos, hogy a halmaz, ami felett elvégezzük az optimalizálást szintén konvex. Ez azért van, mert az ismeretlen változóink a ω vektor, és a b szám, de ezekben a korlát lineáris. Lineáris egyenes pedig konvex.

Állítás. *Az SVM (minimum) feladat megoldása mindig globális minimumot határoz meg.*

Bizonyítás. Korábban már láttuk, hogy egy ilyen optimalizációnak mindig létezik optimális megoldása, tehát mindig van legalább egy lokális optimuma a feladatnak. Az SVM (minimum) feladata (4.2.4) a fentiek alapján konvex programozás, ugyanis konvex halmaz felett konvex függvényt minimalizálunk. Konvex minimum programozási feladatoknak azonban amennyiben létezik lokális minumuma, az globális minimum egyben, Tikhomirov (1996) alapján. □

Észrevétel. Ez az állítás igaz volta teszi az SVM-et az egyik legerősebb Machine Learning technikává. Gyakorlatilag az összes többi gépi tanulási applikáció a lokális minimumok problémájával küzd, azokba előszeretettel konvergál (például a neurális hálók, vagy a k-legközelebbi szomszéd módszere is ilyen). Ez egy nagyon fontos, és nagyon erős érv a Support Vector Machine mellett. Többek között emiatt van az, hogy az SVM, mint modellépítési eszköz ritkán illeszt túl, ellentétben az említett módszerekkel.

A feladat megoldásához használjuk fel a Lagrange-multiplikátor módszert. Erre két okunk is van. Az első, hogy így a korlátok sokkal könnyebben kezelhetőek lesznek. A második, hogy ebben az átírásban a mintánk vektorai csak és kizárólag skaláris szorzatos alakban fognak szerepelni, ami lényeges tulajdonság lesz majd mikor kiterjeszjük az SVM használhatóságát nem csak lineárisan szeparálható esetekre.

Egy kis matematikai trükköt is felhaszálunk: a célfüggvényt $\frac{1}{2}\|\omega\|^2$ alakban minimalizáljuk. A Lagrange-függvény (α jelöli a Lagrange-multiplikátorokat):

$$\mathscr{L}(\omega,b,\alpha)=\frac{1}{2}\|\omega\|^2-\sum_{i=1}^{p}\alpha_i d_i(\omega^T\cdot x_i+b-1)=\frac{1}{2}\|\omega\|^2-\sum_{i=1}^{p}\alpha_i d_i(\omega^T\cdot x_i+b)+\sum_{i=1}^{p}\alpha_i$$

[7]Ortonormált $\{x_1,x_2,\dots,x_j\}$ bázisban ugyanis az euklideszi norma $\|\omega\|^2=<\omega,\omega>=\sum_{j=1}^{p}\chi_j^2$, ahol χ_ja j-edik ortonormált bázisvektorra vonatkozó koordinátája ω-nek. Ezt a több dimenziós Pitagorasz-tételnek nevezik hiszen két dimenzióban pont azt jelenti, hogy a derékszögű háromszögre érvényes az $a^2+b^2=c^2$ híres egyenlet, persze a megfelelő jelölésekkel.

Az ebből származtatható úgynevezett Karash-Kuhn-Tucker feltételek a következők:

$$\begin{bmatrix} 1. & \frac{\partial \mathscr{L}}{\partial \omega} = \omega - \sum_{i=1}^{p} \alpha_i d_i x_i = 0 & \\ 2. & \frac{\partial \mathscr{L}}{\partial b} = \sum_{i=1}^{p} \alpha_i d_i = 0 & \\ 3. & d_i(\omega^T \cdot x_i + b) - 1 \geq 0 & i = 1, 2, \ldots, p \\ 4. & \alpha_i \geq 0 & \forall i \\ 5. & \alpha_i(d_i(\omega^T \cdot x_i + b) - 1) = 0 & \forall i \end{bmatrix} \tag{4.2.5}$$

A Karash-Kuhn-Tucker feltételeknek eleget kell tennie az optimalizációs feladat (4.2.4) megoldásának. Továbbá az SVM feladat konvexitása miatt, a KKT feltételek szükséges és elégséges feltételek az ω, *b*, α trió meghatározásához, így megoldani az SMV feladatot egyet tesz ezen feltételek megoldásával. Emiatt lehet meghatározni a Lagrange-szorzókat a duál-módszerrel. Az optimalizációs feltételek alapján $\omega = \sum_{i=1}^{p} \alpha_i d_i x_i$, és $\sum_{i=1}^{p} \alpha_i d_i = 0$. Ezt írjuk vissza a Lagrange-függvénybe, majd egyszerűsítsünk:

$$\mathscr{L}(\omega, b, \alpha) = \sum_{i=1}^{p} \alpha_i - \frac{1}{2} \sum_{i=1}^{p} \sum_{i=1}^{p} d_i d_j \alpha_i \alpha_j x_i^T \cdot x_j - b \sum_{i=1}^{p} \alpha_i d_i = \sum_{i=1}^{p} \alpha_i - \frac{1}{2} \sum_{i=1}^{p} \sum_{i=1}^{p} d_i d_j \alpha_i \alpha_j x_i^T \cdot x_j$$

De az előbbi miatt az utolsó tag zéró, így a duális feladat:

$$\begin{cases} max & W(\alpha) = \sum_{i=1}^{p} \alpha_i - \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} d_i d_j \alpha_i \alpha_j < x_i, x_j > \\ & \alpha_i \geq 0 \\ feltéve, hogy & \\ & \sum_{i=1}^{p} \alpha_i d_i = 0 \end{cases} \tag{4.2.6}$$

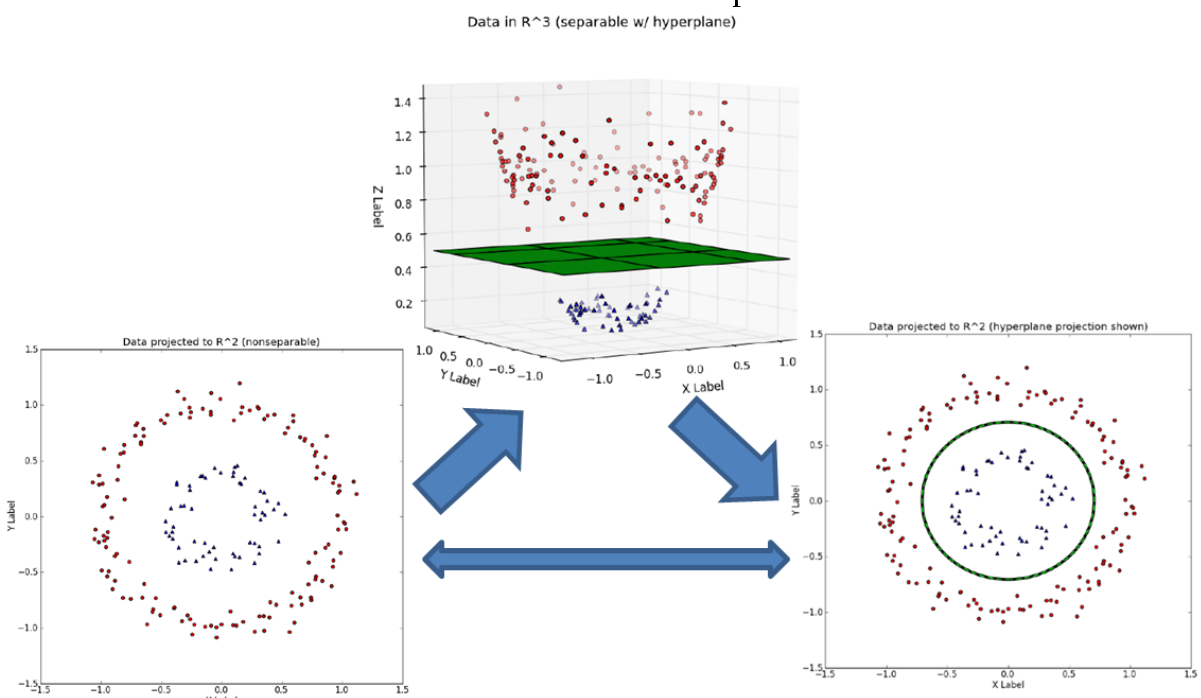
Két dolgot kell most nagyon észbe vésnünk. A legfontosabb dolog az az, hogy a Lagrange-szorzókból rengeteg a nulla értéket fogja felvenni. Csak azoknál a vektoroknál lesz nem nulla az értéke, amik legközelebb vannak az optimális hipersíkhoz, amik azonban éppen a support vektorok! Így adhatnánk másik definicíót is a support vektoroknak: azon vektorok a mintából, amelyekhez tartozó Lagrange-multiplikátor nem nulla. A másik fontos dolog az az, hogy a ω értéke explicite „kijön” a primál feladatból, de hozzátartozó *b* nem, habár kiszámolható lenne az 5-ös KKT feltételből, a support vektorokat tekintve.

Mielőtt tovább mennénk, vessünk egy pillantást az 4.2.5 egyenletrendszer első sorára, ami megadja a ω optimális értékét az optimális α függvényében: $\omega_{opt} = \sum_{i=1}^{p} \alpha_{i_{opt}} d_i x_i$. Tételezzük fel, hogy illesztettünk egy modellt egy tanító halmazra, és most szeretnénk megjósolni egy új *x* megfigyelés kategóriáját. Ehhez a döntési szabályunk (4.2.1) értelmében ki kéne számítanunk a $\omega^T x + b$ értékét, és csak akkor értékelni x-et egyes kategóriába, ha ez az érték pozitív. Felhasználva azonban az összefüggést ez az alábbi formába írható: $\omega^T x + b = (\sum_{i=1}^{p} \alpha_i d_i x_i)^T \cdot x + b = \sum_{i=1}^{p} \alpha_i d_i < x_i, x > +b$. Így, ha megtaláltuk az optimális α_i-ket, amik *mind nullák, kivéve pár elemet, a support vektorokat*, akkor már csak ki kell számolnunk egy mennyiséget, ami egy dologtól függ: a support vektorok és az új elem belső szorzatától. Ez rengeteg memóriát és időt spórolhat meg nekünk az algoritmusunk futásakor, így a duális feladat megoldása célravezető. Következőnek ezt a kiváló tulajdonságot fogjuk kihasználni a lineáris szeparálhatóság feloldásánál.

4.2.2. A bemeneti térben lineárisan nem szeparálható minta

A valóságban a legtöbb statisztikai adathalmaz nem fog lineárisan szeparálható mintát alkotni. Azonban ekkor mindig létezik olyan ϕ transzformáció, amelyet elvégezve a mintaelemeken, már egy nagyobb dimenziós, de lineárisan szeparálható mintát kapunk. Az ilyen tulajdonságú ϕ függvényt bázisfüggvénynek hívjuk, és azt a teret, amiben lineárisan elválaszthatóvá válik a minta, *jellemző* térnek nevezzük. Érzékeltetés képpen, vessünk egy pillantást a 4.2.2 ábrára.

4.2.2. ábra. Nem lineáris szeparalás



Forrás: http://www.eric-kim.net.

A bal oldalon lévő két dimenziós mintán triviálisan látszik, hogy struktúrálisan rendeződik, csak éppen nem abban a lineáris formában, amit keresnénk. Így véve a minta $\phi : \mathbb{R}^2 \to \mathbb{R}^3$, $\phi(x_1,x_2) = \begin{bmatrix} x_1 & x_2 & x_1^2+x_2^2 \end{bmatrix}$ bázisfüggvény általi képét, azt a jellemző térbe transzformátuk (középső kép). A jellemző térben létezik maximális marginnal rendelkező hipersík, így itt az SVM feladatnak van megoldása. Véve a maximális marginnal rendelkező hipersík pontjain a ϕ^{-1} transzformációt, az pontosan a keresett kör alakot fogja nekünk kimetszeni (jobb oldali kép)

Általános esetben egy ismeretlen elem kategóriáját a jellemző térben állapítjuk meg, és az nem más, mint $d(x) = sgn(\omega^T \phi(x) + b) = sgn((\sum_{i=1}^{p} \alpha_i d_i \phi(x_i))^T \cdot x + b) = sgn(\sum_{i=1}^{p} \alpha_i d_i < \phi(x_i), \phi(x) > +b)$.[8]

4.2.2.1. A kernel-trükk

4. Definíció. Azt mondjuk, hogy egy függvény *kernel* típusú, ha igaz rá, hogy $K : \mathbb{R}^2 \to \mathbb{R}$, $K(x,y) = < \phi(x), \phi(y) >$, ahol ϕa bázisfüggvényt jelöli.

Ismerve a bázisfüggvényt, egyszerűen ki tudjuk számolni a kernelt, ha vesszük a belső szorzatukat a jellemző térbeli elemeknek. Ami azonban érdekessé teszi a dolgot az az, hogy sokszor sokkal egyszerűbb kiszámolni $K(x,y)$-t (tehát kevesebb az idő- és memóriaigénye), mint $\phi(x)$-et, $\phi(y)$-t, ugyanis ϕ dimenzionalitása nagyon nagy, akár végtelen is lehet! Talán ezt a legnehezebb megérteni, ezért nézzünk most egy példát. Amikor hasonló alakot ölt a mintánk, mint a 4.2.2 ábrán látott példa, akkor általában polinomiális transzformációval eljuttathatjuk azt a teret a jellemző térbe. Nézzük a négyzetes esetet. Legyen két megfigyelésvektor a térben $a = (a_1, a_2, \ldots, a_n)$ *és* $b = (b_1, b_2, \ldots, b_n)$. Ezekhez rendeljük hozzá azt a vektort, ami rendelkezik az összes elemmel, annak négyzetével, és az összes kétszeres szorzatával is:[9]

$$\phi(a) = (1, \sqrt{2}a_1, .., \sqrt{2}a_n, a_1^2, .., a_n^2, \sqrt{2}a_n a_{n-1}, ..\sqrt{2}a_n a_1, \sqrt{2}a_{n-1}a_{n-2}, .., \sqrt{2}a_{n-1}a_1, ..\sqrt{2}a_2 a_1)$$
$$\phi(b) = (1, \sqrt{2}b_1, .., \sqrt{2}b_n, b_1^2, .., b_n^2, \sqrt{2}b_n b_{n-1}, .., \sqrt{2}b_n b_1, \sqrt{2}b_{n-1}b_{n-2}, .., \sqrt{2}b_{n-1}b_1, ..\sqrt{2}b_2 b_1)$$

Ez már most nagyon félelmetesen néz ki. Elvégezve a $< \phi(a), \phi(b) >$ skaláris belső szorzást azt kapjuk, hogy $< \phi(a), \phi(b) >= \phi(a)^T \cdot \phi(b) = 1 + \sum_j 2a_j b_j + \sum_j a_j^2 b_j^2 + \sum_j \sum_{k>j} 2a_j a_k b_j b_k + \ldots$ Ezt egy sok magyarázóváltozóval rendelkező tanító halmazon nagyon nehéz kiszámolni, ugyanis a két jellemző térbeli vektor belső szorzatát $O(n^2)$ hosszan tudjuk kiszámolni (n^2-el arányos a műveletek számossága). Azonban algebrai átalakításokat végezve, azt kapjuk, hogy $K(a,b) =<$

[8] Az sgn függvény itt az előjelefüggvényt jelöli.
[9] A $\sqrt{2}$-es szorzó csak kényelmi okokból szerepel.

$\phi(a),\phi(b)>= (1+\sum_j a_j b_j)^2$. Így a polinomiális kernel függvényünk ki tudja számolni a belső szorzatot a jellemző térben mindösszesen O(n) számolással! Tehát ahelyett, hogy a két vektort először transzformálnánk a jellemző térbe, és ott vennénk a belső szorzatot, egyszerűen megkaphatjuk az értéket egy könnyű számolással, az eredeti bementi tér adatait felhasználva. Általánosságban ezt a szakirodalom a kernel-trükknek hívja. Azért jogos az elnevezés, mert a kernel függvénnyel egy nagyon nagy dimenziós térbe rakjuk át a problémát (szerepelhet akár a változók összes hatványa, logaritumusa, keresztszorzata, minden amit el lehet képzelni), ami normális körülmények között nagyon nehézkés lenne, vagy nem is működne a nagy dimenzió miatt[10]. A kernel implicit módon definiálja azt a nagydimenziós teret, de úgy, hogy közben ő maga kézben tartható.

Mi lehet kernel? Hogyan tudjuk megállapítani, hogy egy függvényhez létezik jellemző térbe menő transzformáció, ami kielégíti a kernel definíciót? Vegyünk két mintaelemet, például $a = (a_1, a_2, \dots, a_n)$ *és* $b = (b_1, b_2, \dots, b_n)$. Jelölje $\bar{K_{ij}} = K(a_i, b_j)$ a kernel függvény mátrixát. A kernel függvény skaláris szorzat, márpedig a skaláris szorzat konjugált lineáris funkcionál, így a valós térben $K(a,b) = K(b,a)$, azaz a K mátrix *szimmetrikus* ($K^* = K$). Ezenkívül még annak is teljesülnie kell, hogy K pozitív szemidefinit. Általánosságban ez a két feltétel nem csak szükséges, de elegendő feltétel is, J Mercer (1909) tétele alapján.

4.1. táblázat. Gyakran használt Kernel-függvények

Név	Képlet	Jellemző tér dimenziója	Hiperparaméterek
Lineáris Kernel	$K(x,y) = x^T \cdot y$	n	-
Polinomiális Kernel	$K(x,y) = (c + x^T \cdot y)^d$	$\begin{cases} \binom{n+d-1}{d} & c = 0 \\ \binom{n+d}{d} & c \neq 0 \end{cases}$	c,d
Gauss/Sugaras Kernel	$K(x,y) = e^{-\frac{\lVert x-y \rVert^2}{2\sigma^2}} = e^{-\gamma \lVert x-y \rVert^2}$	∞	σ vagy γ
Szigmoid Kernel	$K(x,y) = tanh(\gamma x^T \cdot y + r)$	∞	γ, r

A táblázatban látható harmadik oszlop az SVM kernelek hiperparamétereit tartalmazza. Ezek azok a paraméterek, amelyek megválasztásában szabadon dönthetünk, és éppen emiatt talán a legrelevánsabb kérdések egyike, ugyanis nagyban befolyásolják a klasszifikálás végeredményét. A gyakorlatban legtöbbet a sugaras kernelt használják, ám minden ilyen elemzés során alaposan át kell gondolni, hogy mégis melyik lehet a legcélravezetőbb metódus.

[10]Gondoljunk olyan regresszóra például, amibe sok száz, vagy sok ezer változót rakunk bele.

4.2.3. Lineárisan „nehezen” szeparálható minták

Vannak esetek, mikor a jellemző tér megtalálása még kerneleken keresztül is nagyon hosszadalmas és nehéz. Ekkor használjuk az úgynevezett „puha margin” SVM-et. Ez annyiban különbözik az eddigiektől, hogy a feladat korlátját (4.2.2) gyengítjük, bevezetve minden elemre egy úgynevezett ξ_i slack-változót:

$$d_i(\overline{\omega^T} \cdot x_i + \bar{b}) \geq 1 - \xi_i,\ x_i \in X,\ i = 1, 2, \ldots, p$$

Ekkor triviálisan minden hipersík megoldás, csak a ξ_i-ket kell megfelelő méretűre állítani. Azonban ezeknek a slack-változóknak a nagyságát büntetjük egy úgynevezett C, azaz cost (költség) paraméterrel. Ekkor a feladat primál alakja:

$$\begin{cases} min: & \|\omega\| + C \sum\limits_{i=1}^{p} \xi_i^k \\ & \\ & d_i(\omega^T \cdot x_i + b) \geq 1 - \xi_i \\ feltéve: & d_i \in \{-1, 1\} \\ & i = 1, 2, \ldots, p \end{cases}$$

A cost érték beállítása tetszőleges, ha $C = 0$, akkor semennyire sem büntetjük a hibákat, és minden hipersík megoldása a feladatnak, ha pedig $C \to \infty$, akkor tendálunk az eredeti „hard margin” SVM feladatához, abban az értelemben, hogy semmiféle félrekategorizálás nem engedhető meg. A k érték beállítása is tetszőleges, bár legtöbbször a $k = 1$ változatot használják, aminek a neve *1-Norm Soft Margin SVM*[11]. Az 1-Norm Soft Margin SVM primál Langrange-függvénye a következő:

$$\mathscr{L}(\omega, b, \alpha, \xi, \gamma) = \frac{1}{2}\|\omega\|^2 + C \sum_{i=1}^{p} \xi_i - \sum_{i=1}^{p} \alpha_i [d_i(\omega^T \cdot x_i + b) - 1 + \xi_i] - \sum_{i=1}^{p} \gamma_i \xi_i$$

miközben $\alpha_i \geq 0,\ \gamma_i \geq 0$. Elvégezve a deriválásokat kapjuk, hogy

$$\begin{bmatrix} 1. & \frac{\partial \mathscr{L}}{\partial \omega} = \omega - \sum\limits_{i=1}^{p} \alpha_i d_i x_i = 0 \\ 2. & \frac{\partial \mathscr{L}}{\partial b} = \sum\limits_{i=1}^{p} \alpha_i d_i = 0 \\ 3. & \frac{\partial \mathscr{L}}{\partial \xi} = C - \alpha - \gamma = 0 \end{bmatrix}$$

[11] http://fourier.eng.hmc.edu/e161/lectures/svm/node7.html alapján. A levezetéseket is ez alapján prezentálom.

A feladat duális függvénye ezek alapján:

$$W(\alpha,\gamma) = \sum_{i=1}^{p} \alpha_i - \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p} d_i d_j \alpha_i \alpha_j < x_i, x_j > - \sum_{i=1}^{p} \gamma_i \xi_i - \sum_{i=1}^{p} \alpha_i \xi_i + C \sum_{i=1}^{p} \xi_i$$

$$W(\alpha,\gamma) = \sum_{i=1}^{p} \alpha_i - \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p} d_i d_j \alpha_i \alpha_j < x_i, x_j >$$

Érdekes módon a feladat duális alakja épp megegyezik a lineáris szeparálható SVM duális feladatával (4.2.6), tekintve, hogy $C = \alpha_i + \gamma_i$. Ez alapján az 1-Norm Soft Margin SVM kvadratikus programozás, így hasonlóan oldható meg, mint a lineárisan szeparálható Hard Margin SVM.

A dolgozatban az 1-Norm Soft Margin SVM sugaras kernel változatát fogom alkalmazni, rácsmenti kereséssel (angolul grid search) meghatározva, valamint ötszörös keresztvalidációval kiválasztva a hiperparaméterek és a cost paraméter értékeit.

4.2.4. Az SVM használatáinak előnyei

Ma már a machine learning, mint lassan önálló tudományterület rengeteg technikával rendelkezik (függelék, 4.1.1. táblázat). A support vector machine-t azonban sok minden kiemeli a többi közül, például[12]:

- Az optimális megoldás mindig létezik, sosincs olyan, hogy az SVM nem tud megoldani egy klasszifikálást.[13]
- Az optimum globális optimum. Ez egy olyan fontos tulajdonság, amivel szinte nem is rendelkezik más módszer.
- A kernelek segítségével nagyon könnyen és gyorsan lehet bonyolult, komplex struktúrával rendelkező mintákat egyszerűsíteni, elemezni.
- Az SVM használatához tetszőleges folytonos változókat használhatunk. Nem kell, hogy a tagok függetlenek legyenek, és semmilyen hibatagnak nem kell normális eloszlásúnak lennie.

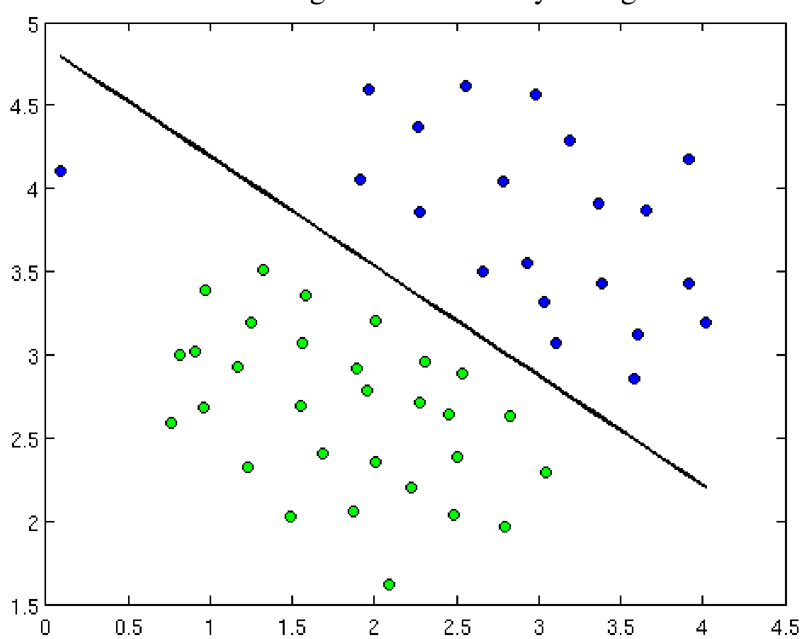
[12] http://svms.org/ alapján.

[13] Az viszont előfordulhat, hogy nem a legjobban, de senkinek sincs mindig jó napja.

4.2.5. Az SVM használatának hátrányai

- A kernel választása tetszőleges. Az optimális kernel megválasztásához maximum csak hüvelykujjszabályokat ismerünk, és leginkább ez igényli a kreativitást a módszer implementálása során. Egy rossz kernel választása, vagy egy rossz hiperparaméter kombináció drasztikusan elronthatja a gép hatékonyságát.
- Az SVM nem tud diszkrét adatokat kezelni, csak folytonosakat. Ez használhatatlanná teszi az SVM-et nagyon sok minta esetében.
- Az SVM nagyon érzékeny az outlierekre. Sajnos egy outlier megléte már a hard margin SVM-et képes tönkretenni (ez látható a 4.2.3 ábrán), míg a soft verzió becslését nagyon torzíthatja. Mindazonáltal már egy kiugró érték is rengeteg későbbi félreklasszifikálást eredményezhet a tesztelő mintánkon.
- Az SVM elvégzése után nem fogjuk tudni, hogy melyik változó hogyan befolyásolja a klasszifikálás végeredményét. Csak egy gyors példa: orvosok nem használhatnak SVM-et, mert nekik tudniuk kell, hogy a páciens megbetegedését mi okozhatja, hogy az ellen például vakcinákat írjanak fel. Ezt úgy szokták mondani, hogy az SVM egy „black-box” módszer.
- Informatikai szemmel nézve az SVM algoritmusa nagyon memóriaigényes, így igénybe veheti a számítógép teljes kapacitását.

4.2.3. ábra. A hard margin SVM érzékeny a kiugró értékékre.



Forrás: http://openclassroom.stanford.edu/

5. fejezet

Az elemzés menete

A dolgozat célja megbecsülni annak a sikerességét, hogy egy tetszőleges vállalat maximum 3 éven belül felvásárlás tárgya lesz. Ehhez szükségünk lesz egy lehetőleg megfelelően nagy elemszámú mintára, amire majd alkalmazzuk az SVM-et. Az adatokat a Budapesti Corvinus Egyetem egyik Bloomberg-termináljáról töltöttem le.

5.1. Adatszerzés

5.1.1. M&A tárgyává vált vállalatok

A szakirodalom néhol kiemeli az országspecifikus változókat, amik esetleg befolyásolhatják a felvásárlás létrejöttét, ilyen például a versenyképesség, vagy a verseny tisztasága az adott országban. Sajnálatos módon azonban a Bloomberg felületen nem található ilyen váltózó, bár valószínűleg megbecsülhető több egyéb változó felhasználásával.[1] Az esetleges szignifikáns változó kihagyása sok bajt okozhat, ezért úgy oldottam fel a problémát, hogy egyetlen országot vizsgáltam a kutatásban, jelen esetben az Egyesült Államokat. A gazdasági válság óta eltelt 10 évet vettem a mintám időperiódusának, tehát adathalmazom 2008. január elseje óta felvásárolt Egyesült Államokbeli cégeket tartalmazza, a 3. fejezetben említett változókkal együtt. Ez pontosan 63 608 datab vállalatot jelent. Természetesen nem létezik tökéletes adathalmaz, így az enyémnek is vannak korlátai. Legnagyobb gyengesége a mintának, hogy számos hiányzó adatot tartalmaz, tehát

[1] Például néhány közgazdasági faktormodellben a közös gazdasági faktort a Vasicek modellel becslik, ami számos más változót használ fel.

gyakorlatilag nincs vállalat, aminek szerepelne az összes változóra értéke[2]. A hiányzó adatok kezeléséről külön szakaszban fogok szólni.

5.1.2. Kontrollcsoport

Kontrollcsoportot megválasztani nehéz. Az említett szakirodalmak közül sokban (példuál Stevens (1973)) úgy válaszották meg a -1-es kategóriájú adathalmazt, hogy adott napra tetszőleges vállalatokat figyeltek meg, amik eddig nem voltak felvásárolva. Nyilvánvalóan ez nem helyes, hiszen olyan vállalatoknak kellene alkotnia a kontrollcsoportot, amiket nemhogy eddig nem vásároltak fel, de *soha nem is fognak*. Az ördög azonban a részletekben rejtőzik: nem tudjuk megállapítani előre, hogy egy céget sose fognak felvásárolni. Ez látszólag egy feloldhatatlan ellentmondás, ezért a TDK dolgozatban nem azt becsülöm meg, hogy egy adott válallatot valaha felvásárolnak-e, hanem azt, hogy maximum 3 éven belül felvásárolják-e. Így a kontroll csoportot olyan módon választottam ki, hogy a korábbi minta időtartamát (2008-2018) visszatoltam az időben 3 évvel (2008-2015), majd letöltöttem a Bloomberg terminálból minden év január elsejére 3000 különböző cégadatot, és megnéztem, hogy az szerepel-e a 5.1.1. alszakaszban már említett cégekkel a következő 3 éven belül. Ha nem, akkor az a bizonyos vállalat a -1-es kategóriát megkapta, ellenkező esetben törölve lett az adatbázisból. A hiányzó adatok kezelése itt is kulcsfontosságú.

5.2. Mintavételezés

A mintavételezés során tartottam magam a Palepu (1986) által javasoltakhoz. Az első azonban nem ő volt, aki felvetette ezt a problémát ezen a területen, hanem Zmijewski (1984). Zmijevski szerint, ha nem abban az arányban válogatjuk össze a felvásárolt, illetve nem felvásárolt cégeket, amilyen az igazi statisztikai aránya, akkor azzal hatalmas torzítást követünk el. Ezt a következő számításokkal illusztrálom, Zmijewski (1984) alapján. Vegyünk egy populációt N vállalattal, ahol N_1 felvásárlás tárgya volt, N_2 pedig nem. A minta álljon n vállalatból, ahol n_1 és n_2 számok rendre a kategóriák számosságai. Legyen $i \in (1,2,\ldots,N)$. Jelölje p annak a valószínűségét, hogy az i-edik vállalat fel lesz vásárolva (ezt szeretnénk megbecsülni), továbbá p' annak a valószínűségét, hogy i a mintában van, és felvásárolják, tehát $p' = P(i\ felvárlás\ tárgya\ lesz|i\ a\ mintában\ van)$. Ekkor a Bayes-tétel alapján:

[2]Ez nem meglepő, mert nagy a modellbe bevont magyarázóváltozók számossága. Ezek megszerzése pedig nagyban függ a cégek reportálási szokásaitól.

$$p' = \frac{P(i\ a\ mintában\ van|i\ f.v) \cdot P(i\ f.v)}{P(i\ a\ mintában\ van|i\ f.v) \cdot P(i\ f.v) + P(i\ a\ mintában\ van|i\ n.f.v) \cdot P(i\ n.f.v)}$$

Random mintavételezés esetén annak a valószínűsége, hogy i a mintában van ugyanannyi, ha fel lesz vásárolva, és akkor is, ha nem, és ezesetben $p = p'$, tehát a p' torzítatlan becslése lesz p-nek. Abban az esetben viszont, ha a mintába nem a statisztikai eloszlás szerint válogatjuk a megfigyeléseket, mindig felülbecsüljük a felvásárlás létrejöttének valószínűségét, az alábbi számolás alapján.

$$p' = \frac{p \cdot \frac{n_1}{N_1}}{p \cdot \frac{n_1}{N_1} + (1-p) \cdot \frac{n_2}{N_2}}$$

$$p' - p = \frac{(\frac{n_1}{N_1} - \frac{n_2}{N_2})p(1-p)}{\frac{n_1}{N_1}p + \frac{n_2}{N_2}(1-p)} > 0\ ha\ N_1 \ll N_2$$

Stevens (1973) például az elemzésében 40-40 fel-, illetve fel nem vásárolt céget vett a mintájába. Tegyük fel, hogy ez egy 1000 elemű populációból származott, és Stevens (1973) beletette az összes felvásárolt céget a mintába. Így a mintavételezési aránya a nem felvásárolt cégeknek 40/960 volt, ami kb 0.042. Vegyünk egy céget, amihez tartozó igaz felvásárlási valószínűség 0.1. A fentiek alapján ekkor a modell által becsült valószínűség ehelyett 0.73 lenne, ami hatalmas hiba lenne. A dolgozatomban ezért törekedtem a helyes statisztikai arány mintán belüli megteremtésére.

5.3. Hiányzó értékek kezelése - Adattisztítás

Talán a legrelevánsabb kérdések egyike az, hogy hogyan kezeljük a hiányzó adatokat egy mintán belül. Ismertetem az legismertebb eljárásokat a következő pontokban, Craig K. Enders (2010): Applied Missing Data Analysis című könyve alapján .

1. Hiányzó adatokkal rendelkező sorok törlése: A legelső és talán legrosszabb módszer az az, hogy minden olyan sort törlünk a mintából, ahol szerepel hiányzó adat. Talán ez nem lenne olyan drasztikus módszer, ha egy-kettő magyarázó változó szerepelne a mintában, és hiányzó értékkel rendelkező sor mondjuk a minta kevesebb, mint 2-3%-át alkotná. Azonban amennyiben ezek egyike nem áll fent, már óriási adatvesztés lép fel. Képzeljük csak el, hogy egy céget, aminek ismerjük 9 ismérvét, csak egyet nem, azt kitöröljük a mintánkból egy ilyen miatt! Ezenkívül amennyiben a hiányzás nem véletlenszerű, úgy a a hiányzó adatok törlésével szisztematikusan torzítjuk a későbbi regressziós eredményeinket.

2. Átlaggal, mediánnal, mozgóátlaggal való helyettesítés: Ezek a legáltalánosabb módszerek a hiányzó adatok kezelésénél, ám ezek is súlyos hibákat hordoznak magukban. Ezekkel a módszerekkel az a baj, hogy úgymond varianciacsökkentőek. Ez egyrészt a különböző modellalkotásoknál jelent problémát, például az egyszerű lineáris regressziónál is a használhatóság egyik fontos feltétele, hogy a magyarázó változó jól szórodjon[3]. Másrészt a mesterségesen kicsi standard hibák hamisan pontossá teszik a modellt. Harmadrészt információvesztéssel is jár: lehet, hogy egy amúgy nagyon jó értékekkel rendelkező vállalathoz így ugyanazt rendeljük, mint egy amúgy nagyon rossz értékekkel rendelkezőhöz. Ekkor tehát a többi változóval való asszociációt nem vettük figyelembe.

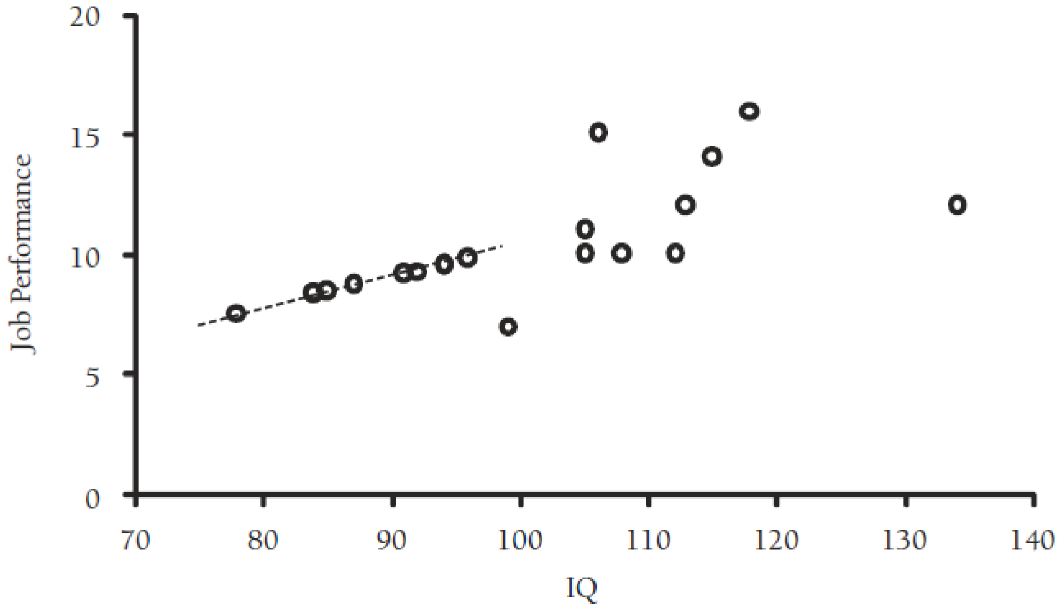
3. A változó historikus eloszlásából való generálás: Hosszabban kifejtve ez azt jelenti, hogy megnézzük, hogy az adott változó, ahol most épp a hiányzó értékünk van, milyen historikus eloszlással rendelkezik, majd abból generálunk egy újat a hiányzó helyére. Ez megoldja a varianciacsökkenés problémáját, de sajnos teljes függetlenséget tételez fel a többi változó és a vizsgált változó között. Mivel ez nem áll fent a legtöbb esetben, ezért ugyanaz a probléma merül fel, mint a 2. pontnál, nevezetesen ez a módszer ugyanolyan valószínűséggel becsül kitűnő értéket alacsony illetve magas értékekkel rendelkező vállalatokhoz.

4. Egyszeres imputálás, regressziós becslések: Ez a fajta becslés úgy pótolja a hiányzó értékeket, hogy regressziót futtat a meglévő magyarázóváltozókra, és az odaillő modellel becsül. Csakúgy, mint a fentebb bemutatott 3 módszer, ez is közkedvelt a statisztikával foglalkozók között. Az alapötlet egyébként nagyszerű: a minták változói gyakran korrelálnak, így van értelme annak, hogy úgymond kölcsön vegyünk információt a többi változótól. Ez a módszer jobb a sima átlaggal való pótlásnál, de sajnos ez is torzít, csak épp az ellenkező módon: ez nem a saját változó eloszlását ferdíti, hanem a többi változó és az adott változó közti asszociációt változtatja meg. Például a lineáris regresszióval való becslés tönkreteszi a változók közti asszociációkat, ugyanis a linearitást jellemző korrelációt növeli. Ezzel párhuzamosan a mintán később végzett modellek hatékonyságát torzítja, például a valós R^2-et felülbecsüli. Ez egyből kítűnik a 5.3.1. képre tekintve.

Így tehát látható, hogy nehéz ezt a problémát megoldani, még haladottabb eszközökkel is. A következőkben Danis (2012) cikke alapján magyarázom az általam használt módszer lényegét. Amikor szeretnénk egy értéket becsülni a hiányzó helyére, akkor 3 kritériumot fogalmazhatunk meg. Mivel a kiegészített adatokkal végzett elemzések révén megbízható és eredményes következtetéseket akarunk levonni a populációról, ezért muszáj megőriznünk a változók eloszlásait és a

[3]Képzeljük el például, hogy úgy akarjuk magyarázni a különböző lakások árát a bennük megtalálható fürdőszobák számával, hogy közben az összes lakásban egy fürdőszoba van. Ekkor ezek számossága nyilván nem hordoz új informácót magában, így használhatatlan.

5.3.1. ábra. A regressziós imputálás hátránya - munkabeli hatékonyság az IQ függvényében



A hiányzó értékeket lineáris regresszióval becsüljük. A változók közti mintabeli lineáris kapcsolat megnő, a korreláció növekszik.
Forrás: Enders (2010), 45. oldal

változók közti asszociációkat (korreláció például). Szeretnénk, hogy a kiegészített változók konfidencia intervalluma fedje a valós értékeket (95%-ban például). Ha a lefedettség pontos, akkor az I. fajú hiba előfordulási valószínűsége is helyes (5% például). Ezenkívűl szeretnénk, hogy konfidencia intervallumok kellően szűkek legyenek, mert ezzel a II. fajú hibák lehetőségei csökkenek. A ma fellelhető legmodernebb és legjobb megoldást úgy nevezik, hogy többszörös imputálás[4], amiről Rubin (1987) írt először, valamint amit Buuren & Groothuis-Oudshoorn (2010) implementáltak az R statisztikai programcsomagban (*mice* package), és ez az, amit én is használni fogok. Ez a technika egy szimuláción és Bayesi-alapokon álló technika, ahol hiányos adatokkal rendelkező mintából $m > 1$ különböző verzió készül el, a hiányzó értékeket valószínűségszámítási apparátussal kiszámolva (például maximum likelihood becsléssel)[5]. Ezután a Rubin (1987) által ismertetett algoritmus szerint kombinálják az eredményeket, ami így várhatóan az lesz, amit mi keresünk. A módszert általában $m = 5$ különböző mátrix elkészítésére korlátozzák, ugyanis az algoritmus számításigényes lehet. Én a TDK dolgozatomban ehelyett 10 különböző imputált mintát használva. A hiányzó értékek kezelése ezen módszerének alkalmazásához való hüvelykujjszabály az, hogy a mintában változónként maximum az adatok 30-40%-a hiányozzon, míg összeségében a minta

[4]angolul multiple imputation

[5]Én ezt úgy képzeltem el, hogy a többszörös imputálás módszere a magyarázóváltozók együttes eloszlásából generál többdimenziós kimeneteleket, és azokat pakolja be a hiányzó értékek helyére. Ezt m>1-szer megtéve, majd az eredményeket úgymond kiátlagolva várhatóan „jó” értékeket fogunk kapni.

10-15%-a legyen hiányos. Az imputálás előtt azonban ellenőriznünk kell pár feltételt a mintával kapcsolatban.

5.3.1. A többszörös imputálás használhatóságának ellenőrzése

Elvégezve az egyáltozós elemzéseket, kidobtam azokat a változókat amiknek a relatív szórása nagyon megközelítette a nullát. Ennek két oka volt. Említettem már a fentebbi felsorolás 2. pontjában, hogy a kis variancia elkerülendő tulajdonság a magyarázó változók esetében, mert nem hordoz ebben az esetben információt a változónk. A második ok pedig az volt, hogy a többszörös imputálás módszere ezen változók esetében gyakorlatilag konstans számokat fog helyettesíteni a hiányzó értékek helyére, ami viszont tovább csökkenti a varianciát. Összességében tehát azokra a változókra nem lehet használni a többszörös imputálást, amelyekben hemzsegnek a hiányzó értékek, illetve amelyek relatív szórása már alapból nagyon kicsi. Az én esetemben, ezért a 63 608 megfigyelést leszűrtem olyan módon, hogy minden megfigyelésnél maximum 5 hiányzó érték legyen (a megmaradt 7-en belül). Így az M&A tárgyává vált vállalatok elemzésben való számossága 2 551 darab lett. Ehhez hozzávettem azokat a vállalatokat a kontrollcsoportból, amelyről megbizonyosodtam, hogy 3 éven belül nem lettek felvásárolva, és hasonlóan megszűrtem őket (ezek számossága 23 450 darab lett).

A többszörös imputációnak van még egy feltétele: a minta típusa MCAR (missing completely at random) vagy MAR (missing at random) kell hogy legyen[6]. Ennek tesztelésére a következő metódust alkalmaztam. Lértehoztam egy dummy mátrixot, amelynek minden eleme 0, vagy 1, attól függően, hogy az adott érték hiányzik-e vagy sem. Majd elkészítettem az oszlopok közti összes lehetséges párosítást, és ellenőriztem az asszociáció erősségét χ^2-teszttel. A feladathoz tartozó kontingencia tábla 2×2-es (NA vagy nem NA), ezért a χ^2-eloszlás szabadságfoka $df = 1$. Ez páronként nézve $\binom{7}{2} = 21$ darab teszt elvégzését jelentette. A 21 kapcsolat közül 1%-os szignifikancia szinten mindösszesen 5 volt független, 17 pedig nem[7]. Ez egyébként nem meglepő, tekintve, hogy a cég adatainak reportálása, jelentése nagyban függ például a cég alkalmazottainak „hátsó” szándékaitól, vagy lezserségétől. Ezt a sajnos rossz eredményt a modell limitációjának (**??**. fejezet) tekintem.

[6]Ha ezek egyike se áll fent, akkor a mintavételezéssel lehetnek problémák. A vállalatok egy specifikus részének hiányzik valamely értéke. Ekkor további elemzéseket kell elvégezni. A vállalatok mely része hanyag a reportálás során?

[7]Ez az eredmény a MCAR-t veti el, a MAR-t nem.

5.1. táblázat. Példa a hiányzó értékek asszociációinak kapcsolatáról - adattisztítás előtt.

	első	második	khi teszt értéke	p érték	független
14	Target Sales to Total Assets	Target EBIT Margin	7.15474546973866	0.007	FALSE
15	Target Sales to Total Assets	Target Dividend Payout Ratio	2.02933765360322	0.154	TRUE
16	Target Sales to Total Assets	Target Current Assets to Total Assets	0.575548694433831	0.448	TRUE
17	Target Sales to Total Assets	Target Cash to Total Assets	0.109443504569818	0.741	TRUE
18	Target Sales to Total Assets	Target Cash Flow per Share	0.702753881958817	0.402	TRUE
19	Target Sales to Total Assets	Target Book Value per Share	0.447063136289827	0.504	TRUE
20	Target Sales to Total Assets	Target Price Earnings Ratio (P/E)	2.58162148776143	0.108	TRUE
21	Target Sales to Total Assets	TV/EBITDA	4.02593026537509	0.045	TRUE
22	Target Return on Total Equity (Including Pref.)	Target Profit Margin	7.47650693854498	0.006	FALSE

5.4. Az outlierek szűrése

Az imputálás elvégzése előtt továbbá előre kiszűrtem az outliereket a tanító halmazomból. A statisztikusokat, gépi tanulással foglalkozó szakembereket nagyon megosztja ez a kérdés, de hosszabb kutatás után az outlierek mellőzése mellett döntöttem. Erre két okom volt. A Support Vector Machine sajnos nagyon érzékeny az outlierekre, és komolyan befolyásolhatja a klasszifikálás végeredményét, a fentebb összefoglaltak alapján. Ezenkívül véleményem szerint az outlierek megléte a többszörös imputálás végeredményét is torzítja, ugyanis nagyítja a varianciáját a változóknak, amivel az imputált adatok hamis extremitásokat vehetnek fel. A kiugró értékektől való megszabadulást vertikális és horizontális outlier teszttel végeztem el[8].

A vertikális outlier-kiszűrési módszer során az úgynevezett winsorizálás metódusát követtem. A winsorizálás során a változókat külön-külön sorba rendeztem, majd a 95% feletti percentilishez tartozó értékeket a 95%-os percentilisnél lévő értékkel helyettesítettem, illetve az első 5%-hoz tartozó megfigyeléseket is az 5%-os megfigyeléssel helyettesítettem. Ezzel információt vesztünk, de tapasztalataim alapján így az SVM kevésbé fog túlilleszteni és összességében hatékonyabban fog rátanulni a mintára. A módszer implementálva van R-ben, a *RobustHD* packageben. A package dokumentációja javasolja a változók úgynevezett robusztus sztenderdizálását, ami lényegében azt jelenti, hogy nem az átlagra kell centrálni a mintát első lépésben, hanem a mediánra. Ez könnyen magyarázható: amennyiben vannak hatalmas outlierek, úgy az átlag torzízott becslése lesz a minta változójának várható értékének. Ekkor sokkal jobb becslést ad a medián, így én is robusztus sztenderdizálást alkalmaztam.

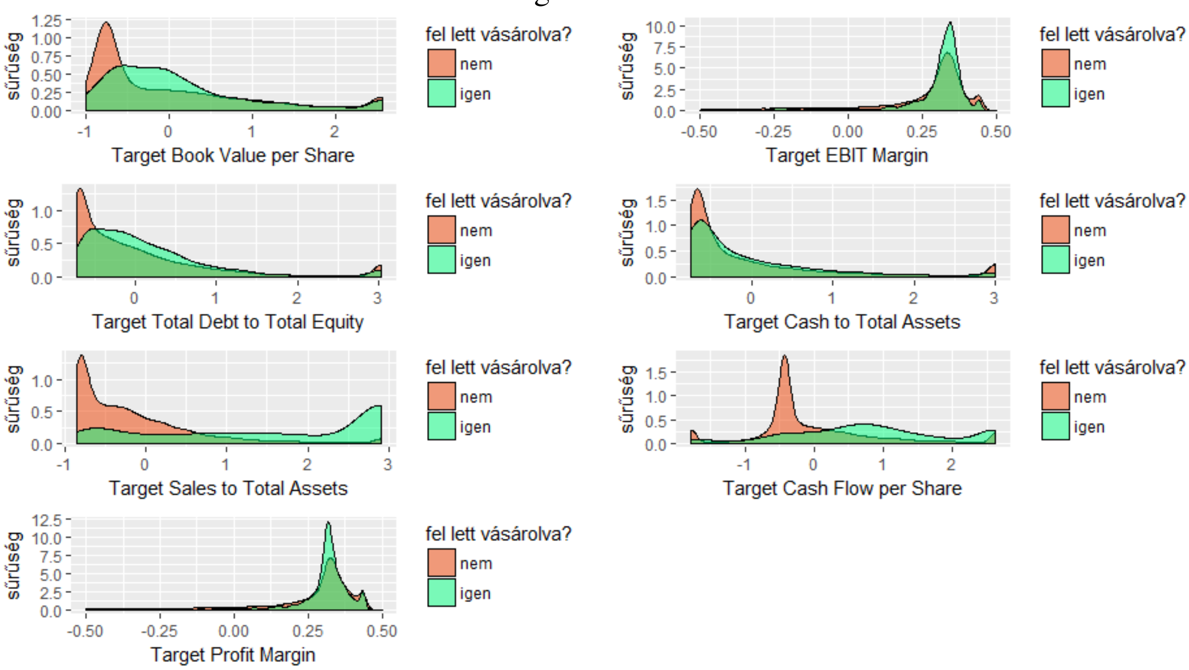
A horizontális outlier detektálás módszert saját ötlet alapján implementáltam. A megfigyelések vektorokat alkotnak a magyarázó változók által kifeszített térben, és mint ilyeneknek, van hosszuk.

[8] Saját elnevezés.

Véve az Euklideszi-metrikát (önmagukkal vett belső szorzatuk gyökét), a megfigyeléseket hossz szerint sorba rendeztem[9]. A legfelső 2.5%, illetve a legalsó 2.5%-ot ebben az esetben eldobtam a mintából. Fontos látni, hogy most nem tehetjük meg azt, amit előbb, mert a hosszakra tekintve nem tudjuk, hogy milyen előjelű változókból származhatott a kérdéses megfigyelés, így nem tudjuk pontosan megmondani, hogy mivel helyettesítsünk. Érdemes megjegyezni továbbá, hogy amennyiben a felvásárolt és nem felvásárolt cégek lineárisan szeparálható mintát alkotnának, úgy a horizontális szűrés elvégzése hatalmas hiba lenne, ugyanis pont a hatalmas hosszbeli különbségek differenciálják a kategóriákat. Azonban ennek az ellenkezőjéről az elemzés során megbizonyosodtam.

Az outlierek eltávolítása után a minta 22 197 nem felvásárolt, és 2 504 felvásárolt vállalatot tartalmazott. Ez a megtisztított minta már alkalmas volt a többszörös imputálás elvégzésére, amiután 10 darab hiányzó értékkel nem rendelkező tanító halmazt kaptam végeredményül. A 10 imputált mátrixnak vettem a pontonkénti átlagát, amin felépíthettem a dolgozat hatáskörébe tartozó két modellt. Az így kapott változók feltételes eloszlásait ábrázolja a 5.4.1 lévő ábra.

5.4.1. ábra. A változók feltételes sűrűségei outlier-szűrés és robusztus sztenderdizálás után



Saját ábra

[9]Természetesen a robusztus sztenderdizálás után

5.5. SVM alkalmazása

A Support Vector Machine alkalmazásához mindenekelőtt sztenderdizálni kell a változóinkat, ugyanis a módszer érzékeny a mértékegységekre. Sztenderdizálás után minden szórásegységben van kifejezve, így az SVM alkalmazható. Az applikáció Min & Lee (2005) cikke alapján történik. A szerzők kiemelik, hogy az SVM-közelítés könnyen alkalmazható pénzügyi problémák megoldására, például hitelbesorolásra, idősoros klasszifikálásra és biztosítási csalás detektálására. A cikk szerint az SVM versenybe tud szállni a hagyományos eszközökkel, és sokszor jobban is teljesít náluk.

A Gauss-féle kernelnek két hiperparamétere van: a szigma (vagy gamma) és a büntetés paraméter (4.1. táblázat). A dolgozatban a szigmával számoltam, amiből a gamma azonban egyszerű számolással megkapható, ugyanis $\gamma = \frac{1}{2\sigma^2}$. A σ és a C paraméterek helytelen kiválasztása komoly problémákat okozhat, ugyanis okozója lehet a túl-, illetve alulillesztésnek is. Másképp szólva, a kutatás eredménye a hiperparaméterek helyes kiválasztásán *teljes mértékében áll vagy bukik*. Ezen ok miatt, nagyon nagy körültekintéssel optimalizáltam a paramétereket, 5 hálós keresztvalidálást végezve. A módszer során a Hsu *et al.* (2003) által tanácsolt rácsmenti keresésből indultam ki, ami a lehetséges paraméter-párok egy durva felbontásán halad át, majd egyenként kiértékelve a modell illeszkedését, kiválasztja a legjobbat. A rácsmenti keresés elnevezés onnan származik, hogy adunk egy lehetséges számhalmazt a C illetve σ értékeknek, majd azok minden párosításán, mint rácspontokon lépdelve megkeressük az optimálisat közülük. A szakirodalom alapján exponenciálisan növekvő sorozatokat érdemes megadni a lehetséges értékeknek, így én a σ értéket a $2^{-6}, 2^{-5}, \ldots, 2^{2}$ számok közé szorítottam, míg a büntető paramétert a $2^{0}, \ldots, 2^{8}$ halmazból kerestem ki (hasonlóan Min & Lee (2005) munkájához, akik SVM-et használtak bankcsőd megtörténésének előrejelzésére).

A modellalkotás során nem feltétlenül érdemes elérni egy nagyon magas tanuló-pontosságot, hiszen az könnyen a túlillesztés proxyja lehet. Mégis, a rácsmenti keresés során ezt maximalizáljuk, mert másképp nem tudnánk optimálisat választani a paraméterek közül. A túlillesztés detektálására azt a szabályt alkalmaztam, hogyha a tartóvektorok száma a mintaelemszám több, mint 10-15%-át teszi ki, akkor igen valószínű a túlillesztés bekövetkezte, bár ez csak egy egyénileg megfogalmazott határ[10].

A validáló mintán elért eredmény egyszersmind e dolgozat végső eredménye, hiszen ez adja meg, hogy milyen pontossággal becsülöm meg a felvásárlás tényét a következő 3 évben, tetszőleges egyesült államokbeli vállalatot tekintve.

[10] A support vektorok a marginhoz legközelebbi pontok. Ha ebből sok van, akkor a margin „kígyózik” a pontok között, nem pedig a minta struktúráját tanulta meg. Ezért intuitív a metrika használata.

5.5.1. Keresztvalidációs hiperparaméter optimalizálás

A keresztvalidáció implementálása remekül megvan oldva az R *caret* nevű csomagjában, így ezt használtam. A lehetséges (σ, C) paraméterek mind a 64 rácspontjának az esetében külön-külön felosztottam a minta 70 %-át alkotó tanító halmazt 5 véletlenszerűen kiválasztott, különálló részre. Minden rácspont mellett elvégeztem a modell kalibrálását 4 darab ötödön, majd az ötödiken teszteltem azt. Ezután másik 4 ötödön tanítottam, hogy a maradék egyen tesztelhessek. Összességében 64 rácspontra külön-külön 5 modellt számítottam ki, ami 320 modellt jelent. Egy (σ, C) páros eredménye így az 5 különböző modell 5 különböző tesztelő halmazán elért hatékonyságnak a számtani átlaga. Ezt elvégeze az összes lehetséges (σ, C) pároson kiválasztottam, *keresztvalidáltam* a legjobban teljesítő hiperparaméter-párt, tehát a 64 átlagolt pontossági értékek közül kiválaszom a maximálisat.

A leírt metódus alapján az ötszörös keresztvalidációs technikával a $(\sigma, C) = (1, 8)$ pár rendelkezett a legjobb eredménnyel, így a validáló halmazon majd ezen hiperparaméterekkel fogok előrejelezni. Ezekkel a paraméterekkel az SVM a tanuló halmazon 95.8878% pontosságú volt. Általános hüvelykujjszabály, hogy a túlillesztés detektálható, ha a support vektorok aránya a teljes mintán belül meghaladja a 15%-ot. Jelen esetben 2 534 tartóvektor volt, amely 14.656%-a az egész tanuló mintának, így a túlillesztés „vádja” éppenhogy, de elvethető.

5.6. Validálás, végső eredmény

A validálás során több út is járható lett volna, például a validáló halmaz külön imputálása, majd az azon való előrejelzés, vagy a hiányzó értékek teljes figyelmen kívűl hagyása, és az azon való validálás stb. A dolgozat írása közben végig az lebegett a szemem előtt, hogyha a Kedves Olvasó rámutat egy tetszőleges cégre, hogy „na most mondd meg a tutit!”, akkor tudjak neki mondani valamit. Az teljesen nyilvánvaló az adatok struktúrájából, hogy gyakorlatilag nincs cég, amelyhez a Bloomberg minden adatot ismerne. A modellem azonban nem tud hiányzó értékkel rendelkező megfigyelésre előrejelezni, hiszen egyrészt teljes mintán tanult, másrészt úgy a magyarázóváltozók által kifeszített térben egy vagy több bázisvektorhoz nem tartozna koefficiens, így viszont az előrejelezendő cég nem illeszhető be a modell keretei közé. Viszont ha egy cégre kell prediktálnom, akkor az is teljesen egyértelmű hogy arról nem tudom megmondani, hogy kiugró értékekkel rendelkezik-e, vagy hogy a hiányzó értékeit mivel lehetne legjobban pótolni. Ezeket a kérdéseket, csak *összefüggéseiben* tudom megválaszolni, a train mintára tekintve!

Tehát az előrejelzés során az új vállalatot hozzácsatolom a tanulómintához (még sztenderdizálás előtt), majd ott elvégzem a winsorizálást. Amennyiben az új vállalat extrém értékekkel rendelkezett, úgy ki lesz cserélve az összes változója az oda tartozó (95%-os vagy 5%-os percentilishez tartozó) vektorral, de amennyiben nem, úgy benne hagyom az új, összecsatolt mintában. A minta fog tartalmazni hiányzó értéket, ezért újra el kell végeznem a többszörös imputálást az egyesített mintán! Ekkor megoldottam mindkét problémát: az új vállallat esetleges kiugró mivoltát, illetve hiányzó értékeinek kezelését. Ezen az új imputálát mátrixon fogom lefuttatni a modellt ezután!

Tehát a validáló halmaz hiányzó értékeinek ezen kezelése, és extremitásának megvizsgálása után az előző pontban épített, optimális hiperparaméterekkel rendelkező Support Vector Machine az 5.2 kontingenciatáblával jellemezhető eredményt érte el a teszt halmazon:

5.2. táblázat. Az SVM előrejelző képessége vállalati felvásárlások alapján

$Valóság \backslash Előrejelzés$	Nem felvásárolt	Felvásárolt
Nem felvásárolt	6566	75
Felvásárolt	312	458

A kontingencia tábla alapján a modell pontossága, tehát helyesen klasszifikált megfigyeléseinek részaránya, és egyben a dolgozat egyik legfontosabb eredménye: $\mathbf{94.778}\%$.

6. fejezet

Arbitrázs stratégia kialakítása

„Although efficient markets people still go around saying there is a "mountain" of evidence supporting their hypothesis, the truth of the matter is that it's a very old mountain that's now eroding rapidly into the sea.”[1]

Robert Haugen

Az absztraktban illetve a bevezetőben a következő kérdést fogalmaztam meg. Lehetséges-e a modell előrejelzésére kereskedési stratégiát építeni, valamint ezzel a stratégiával a kockázattal korrigált piaci hozamnál szignifikánsan magasabb prémiumot elérni? Erre a kérdésre igenlő a válasz. A stratégia mindig egy-egy cégre vonatkozik. Ha egyszerre számos választási lehetőségünk van, akkor nagyban az egyéni preferenciákon dől el a választás kérdése.

6.1. Az M&A-n nyerhető várható hozamok

Ezen dolgozat elején említettem, hogy a felvásárlási ügylet többletet termel, így a felvásárló, és a felvásárolt vállalatok részvényárfolyama szimmetrikusan emelkedik egy ilyen bejelentés hatására. Az információk időben elnyúlva épülnek be az árakba, így a felvásárlás előtti hónapokban már elkezdenek emelkedni az árfolyamok, egészen a deal bejelentése utáni hónapokig. Barnes (1996) szerint egy felvásárolt cég átlagosan 31.25%-ot értékelődik fel a bejelentés napjához képest

[1] „Habár a hatékony piac mellett érvelők még mindig azt állítják, hogy a hipotézisüket „hegynyi” bizonyíték támasztja alá, az igazság az, hogy ez a hegy már nagyon öreg és gyorsan omlik a tengerbe.” - saját fordítás. Forrás: Haugen (1999), 1. fejezet, Bevezetés, 2. oldal

2 hónappal korábbi naptól a bejelentés utáni 3. hónapig. Crawford & Lechner (1996) szerint a bejelentés előtti 50. naptól a target vállalat kivezetéséig várhatóan 39%-os emelkedés figyelhető meg. Dolgozatomban azonban ragaszkodom a Palepu (1986) által közölt számokhoz, ami 20.98% a felvásárlás, és -1.62% a felvásárlás félrekategorizálása esetén. Ezzel azt szeretném elérni, hogy egy konzervatív becslést adjak az elérhető nyereségre: ha a Palepu (1986) féle számok alapján is jövedelmező a stratégia, akkor szinte biztosak lehetünk benne, hogy a valóságban is használható az itt közölt módszer.

6.2. A stratégia

$$stratégia: \begin{cases} Vegyél\ részvényt! & , ha\ a\ gép\ felvásárlást\ predikál. \\ Ne\ vegyél\ részvényt! & , ha\ a\ gép\ nem\ predikál\ felvásárlást. \end{cases}$$

A stratégia mögötti intuíció az az, hogy az előrejelzés majdnem biztosan megmondja, mely vállalatokat nem fognak megvenni a következő 3 évben. A modell a kontingencia tábla alapján a teszt halmazon $6566+312=6878$ vállalatra predikált nem felvásárlást, és ennek $\frac{6566}{6878}\cdot 100 = 95.46\%$-a helyes is volt. Tehát ha nem predikál felvásárlást a modell, akkor ne fogadjunk az ellenkezőjére, hiszen az 95.46% valószínűséggel negatív 1.62%-nyi többlethozamot fog nekünk eredményezni, 4.54% valószínűséggel pedig pozitív 20.98%-ot, ami várhatóan $(0.0454\cdot 20.98\% - 0.9546\cdot 1.62\% = -0.593\%)$ veszteséges üzlet lesz - Palepu (1986) cikke és a bemutatott számítások alapján. Azonban amennyiben a modell felvásárlást prediktál, akkor az esetek $\frac{458}{458+75}\cdot 100 = 85.92\%$-ban ez az előrejelzés helyes volt. Így a vegyél részvényt! szignál alapján 85.92% valószínűséggel pozitív többlethozamunk termelődik, számszerűen 20.98% (ismét Palepu alapján), 14.08% eséllyel pedig nem jön be, és ekkor negatív többletünk termelődik: -1.62%. A várható többlethozam így felülmúlja a várható büntetőhozamot, tehát valószínűségi arbitrázst alakítottunk ki: $0.8592\cdot 20.98\% - 0.1408\cdot 1.62\% = 17.79\%$. A stratégia használatakor $\frac{312}{312+458}\cdot 100 = 40.51\%$-nyi M&A-t elszalasztunk, de ez a portfóliónkban veszteséget nem okoz, hiszen a Ne vegyél részvényt! szignált mi nem fogadjuk meg. Az eredményt viszonyítsuk össze a következővel. A kutatás hatáskörébe tartozó amerikai vállalatok alapján vegyünk tetszőleges 1000, a Bloombergen megtalálható amerikai vállalatot. Ezek mindegyikébe fektessünk egy egységnyi dollárt. A tévesztési mátrix alapján a vállalatok 7.75%-ára a gépünk Vegyél részvényt!, míg 92.5%-ára Ne vegyél részvényt! szignált küld. Ekkor 75 cégen átlagosan nyerünk 17.79% százalékot, 925 cégen pedig átlagosan vesztünk 0.593%-ot. Így az a passzív stratégia, hogy vegyünk annyi részvényt amennyit csak tudunk, várhatóan 0.785%-os többlethozamot termel[2] . Ehhez képest a gépünk a portfóliónkban

[2]Ez nem meglepő, a részvények általában pozitív trendet követnek. Erre épül a híres Buy and Hold stratégia.

17.79%-os többlethozamot generál, ami kb. 23-szor annyi, mint a passzív stratégia várható kifizetése.

7. fejezet

Limitációk

A modellemnek, mint minden statisztikai modellnek sajnos megvannak a maga gondjai, limitációi. Itt összefoglalom azokat, amikre rájöttem a modell konstruálása közben.

1. Sérül a hiányzó értékek függetlenségének elve. Ezt a bemutatott számolás alapján állíthatom. Ennek a feloldása nem triviális, talán főkomponenselemzéssel és varimax rotációval meg lehetne oldani a függetlenség elfogadható teljesülését. Ez további elemzéseket von maga után.

2. A hiányzó értékek információtartalmától való megszabadulás. Említettem a dolgozatban, hogy sok olyan változó volt a mintában, amik sokkal inkább hiányoztak akkor, ha a felvásárlás megtörtént, mint mikor nem (például P/E). Ez alapján a felvásárlás tényét az adott változó hiányzása képes előrejelezni, vagy legalábbis a hiányzásra képes lenne rátanulni az SVM. Egyelőre nem világos számomra, hogy ezt hogy oldhatnám meg. Ez további elemzést igényel.

3. A szigma és a büntető paraméterek halmazának korlátozása. Sajnos a C és a σ paramétereket mind-mind egy 8 elemű halmazból választottam ki. A rácsmenti keresés és az ötszörös keresztvalidáció azonban olyan lassúvá teszi az algoritmust, hogy a számítógépi kapacitásom nem tudott bővebb halmazt megvizsgálni. Szakirodalom alapján (Min & Lee (2005)) próbáltam a lehető legjobb számkombinációkat kiválasztani, de ez nyilván mintafüggő. Az általam kapott eredmény mellett szól, hogy belsőponti paraméterkombináció, tehát a szomszédos 4 rácspontban kisebb a modell hatásfoka. A paraméterkombináció optimumának globális voltát, tehát hogy amennyiben mindkét lehetőséghalmazt kibővíteném két oldalra, akkor is ugyanide konvergálna-e az eredmnény vagy sem, nem tudtam megvizsgálni.

4. A végső eredményül kapott kontingencia mátrix részben véletlenfüggő. Igyekeztem a lehető legjobban csökkenteni az eredményben az imputálás éppeni realizációját, vagy a tanuló halmaz véletlenségének hatását, de ezt sose lehet megtenni biztosan. Ezt a limitációt úgy lehetne feloldani, hogy egy Monte-Carlo szimuláció során 1000-5000 kontingencia mátrixot számítok ki, majd ezeknek végül veszem a pontonkénti átlagát. Tekintve, hogy már egy ilyen mátrix kiszámítása is hatalmas mennyiségű számítást igényel, ezt én nem tudtam megtenni. Ennek ellenére megérzésem alapján az eredményül kapott 17.79% benne van a várható többlethozam 97-98%-os konfidencia intervallumában (ezt a különböző realizációk megvizsgálása miatt állítom).

5. A modell tesztelésének fontos limitációja, hogy a tesztmintán belül a felvásárolt vállalatok aránya a saját mintánkat tekintve reprezentatívnak tekinthető, ugyanakkor nem rendelkezünk pontos adatokkal a sokasági statisztikákról. Már az összes, potenciálisan felvásárlás tárgyává váló vállalatok halmazának számosságának meghatározása is nehézkes, hiszen a nagyon kis (például KKV-hez hasonló vállalatok) és nagyon nagy vállalatok (a gazdasági versenyhivatal nyomására) szinte kizárt, hogy felvásárlásra kerüljenek. TDK dolgozatomban a felvásárolt vállalatok tesztmintabeli aránya 11.6%, összehasonlításképpen Harris *et al.* (1982) 4.8%-os arányt, Palepu (1986) pedig 2.69%-os arányt alkalmaz. Amennyiben a felvásárolt vállalatok felülreprezentáltak a tesztmintában, akkor a modell felülbecsli a sikeresen kategorizált megfigyelések arányát, így az eredmények értelmezésénél érdemes ezt a limitációt észben tartani.

6. A modell használata során feltételezem, hogy a sikeres, illetve nem létrejövő felvásárlások a magyarázóváltozók terében időben azonosan strukturálódnak. Ez ekvivalens azzal a limitációval, hogy feltételeztük, hogy a szeparáló hipersík, tehát az SVM hiperparaméterei időben állandóak. Ezt úgy lehet megoldani, hogy minden egyes létrejövő felvásárlást, illetve nem felvásárlást hozzáillesztünk a tanulómintához, majd újrakonfiguráljuk a modellt. Ezzel a limitációt megoldottuk. Ez a validáló halmazban lévő 7410 cégre azonban nagyon számításigényes lett volna, így ezt nem tettem meg.

7. A dolgozat eredményéül kapott hozamszámítás nem tartalmazza a különböző hibák súlyát.

8. A többszörös imputálás módszere úgy használható, hogy a létrejövő mintákon külön-külön elvégezzük a modellalkotást, majd az eredményeket Rubin (1987) metódusa alapján összegezzük. A hatalmas számításigény miatt azonban én a mintákat pontonként átlagoltam, majd az átlagos imputált mintán alkottam meg a modellt. Ez a módszer így már nem a többszörös imputálás, hanem annak egy gyengített verziója. Az eredmények erős szignifikanciája azonban arra enged következtetni, hogy a kereskedésen nyerhető periódusonkénti vagyonnövek-

mény a dolgozatban leírtnál csak *nagyobb* lehet. Így a szakdolgozatom megőrzi konzervatív mivoltát.

8. fejezet

Összefoglalás

TDK dolgozatom célja a Support Vector Machine (SVM) módszerének alkalmazása vállalatok klasszifikációjára aszerint, hogy felvásárlásra érdemesek-e. A múltban illesztett statisztikai modellek zömében a diszkriminancia-analízis és a logisztikus regresszió módszereit alkalmazták, annak ellenére, hogy az SVM módszere ezen a területen is ígéretesnek tűnik. A korábbi, vállalati felvásárlások előrejelzését célul kitűző statisztikai munkák jelentős része 50%-os arányban határozta meg mind a tanuló, mind a tesztmintán belül a felvásárolt vállalatok arányát, viszont e cikkek kritikájából kiderül, hogy ez a felvásárlás valószínűségének a modell általi felülbecslését vonja magával, így ügyeltem a felvásárolt vállalatok mintán belüli reálisabb arányának megválasztására. A vizsgált szakirodalmakban továbbá nem volt lehetőség a felvásárlás előrejelzésére stratégiát építeni.

A pénzügyi szakirodalom alapján választottam ki a modellbe bevonandó magyarázó változókat, amik az eszközarányos árbevétel, az összértékesítés/összes eszköz, EBIT/árbevétel, a részvényegységre jutó cash flow (melyek a menedzsment hatékonyságát hivatottak operacionalizálni), az összes adósság/saját tőke (ami az optimális tőkeszerkezethez való közelséggel függ össze), a készpénz/összes eszköz (ami az erőforrásbőséget fejezi ki) és a könyv szerinti érték/piaci kapitalizáció (ami a felvásárlandó vállalat alulértékeltségét mutatja) voltak. Az elemzés során használt minta elemszáma lényegesen nagyobb volt, mint a felhasznált szakirodalmakban, ami a rendelkezésre álló Bloomberg adatbázisnak köszönhető. A korábbi munkák legfeljebb 1 100-1 200 körüli mintaelemszámot alkalmaznak, míg az e dolgozatban használt minta 23 450 fel nem vásárolt és 2 551 felvásárolt vállalatot tartalmazott. A jelentős mennyiségű hiányzó információ problémáját a többszörös imputálás módszerével oldottam fel, ami a lehetséges opciók közül a legkedvezőbb volt, tekintve, hogy a módszer figyelembe veszi és megőrzi az egyes változók összefüggőségi struktúráját.

A jelentős mennyiségű hiányzó adat problémáját a többszörös imputálás módszerével oldottam fel, ami a lehetséges opciók közül a legkedvezőbb, tekintve, hogy figyelembe veszi az egyes változók összefüggőségi strukturáját. Az optimális SVM hiperparaméter kombinációt rácsmenti kereséssel és ötszörös keresztvalidációval állapítottam meg. Az ezekhez tartozó optimális modell pontossága 94.78%-os lett, a tartóvektorok 14.66%-os tanuló halmazbeli aránya mellett, ami mutatja, hogy a modell még éppen a túlillesztés szokásos 15%-os kritériuma alá esik. A modell tévesztési mátrixa és Palepu (1986) felvásárolt és fel nem vásárolt vállalatok többlethozamaira vonatkozó kutatása lehetőséget adott arra, hogy megvizsgáljam, a modellünk segítségével lehet-e egy alkalmas benchmarkhoz képest (esetünkben Palepu (1986) cikkének nyomán a hasonló bétával rendelkező cégek várható hozamához képest) többlethozamot elérni.

Tekintve, hogy modellem a felvásároltnak előrejelzett vállalatokat 85.9%-ban kategorizálta helyesen a tesztmintán belül, és Palepu (1986) nyomán az ilyen vállalatok többlethozama 20.98% (a felvásárlás tárgyává nem váló vállalatok többlethozama pedig -1.62%), így a modell szerint valóban lehetőség van a vállalati felvásárlásokra való fogadással statisztikai arbitrázsra, ami ennek a piacnak a hatékonyságának közepesen erős változatát tagadja. Fontos ugyanakkor a modell hiányosságait szem előtt tartani az eredmények értelmezésekor, valamint a módszertani limitációk mellett megjegyezni, hogy amennyiben a vállalati felvásárlások sokasági aránya a tesztmintán belüli arányhoz képest alacsonyabb, a modell pontosságát felülbecsüljük.

Irodalomjegyzék

BARNES, PAUL. 1996. The regulation of insider dealing in the UK: some empirical evidence concerning share prices, merger bids and bidders' advising merchant banks. *Applied Financial Economics*, **6**(4), 383–391.

BELKAOUI. 1978. Financial ratios as predictors of Canadian takeovers. *Journal of Business Finance & Accounting*, **5**(1), 93–107.

BODIE, ZVI, KANE, ALEX, & MARCUS, ALAN J. 2013. *Investments, 10th Edition*. McGraw-Hill Education.

BUUREN, S VAN, & GROOTHUIS-OUDSHOORN, KARIN. 2010. mice: Multivariate imputation by chained equations in R. *Journal of statistical software*, 1–68.

CRAWFORD, DEAN, & LECHNER, THOMAS A. 1996. Takeover premiums and anticipated merger gains in the US market for corporate control. *Journal of Business Finance & Accounting*, **23**(5-6), 807–829.

DANIS, ILDIKÓ. 2012. Az adathelyettesítés modern technikája-„Multiple Imputation (MI)". *Alkalmazott Pszichológia*, 65–70.

DIETRICH, J KIMBALL, & SORENSEN, ERIC. 1984. An application of logit analysis to prediction of merger targets. *Journal of Business Research*, **12**(3), 393–402.

ENDERS, CRAIG K. 2010. *Applied Missing Data Analysis (Methodology in the Social Sciences)*. The Guilford Press.

FOGELBERG, GRAEME, LAURENT, CLINTON R, & MCCORKINDALE, DEREK G. 1978. The usefulness of published financial data for assessing the takeover vulnerability of companies in New Zealand. *New Zealand Economic Papers*, **12**(1), 184–206.

HARRIS, ROBERT S, STEWART, JOHN F, GUILKEY, DAVID K, & CARLETON, WILLARD T. 1982. Characteristics of acquired firms: fixed and random coefficients probit analyses. *Southern Economic Journal*, 164–184.

HAUGEN, ROBERT A. 1999. *The inefficient stock market: What pays off and why*. Prentice Hall.

HSU, CHIH-WEI, CHANG, CHIH-CHUNG, LIN, CHIH-JEN, *et al.* 2003. A practical guide to support vector classification.

J MERCER, BA. 1909. XVI. Functions of positive and negative type, and their connection the theory of integral equations. *Phil. Trans. R. Soc. Lond. A*, **209**(441-458), 415–446.

JENSEN, MICHAEL C, & RUBACK, RICHARD S. 1983. The market for corporate control: The scientific evidence. *Journal of Financial economics*, **11**(1-4), 5–50.

LEWELLEN, WILBUR G. 1971. A pure financial rationale for the conglomerate merger. *The Journal of Finance*, **26**(2), 521–537.

MIN, JAE H, & LEE, YOUNG-CHAN. 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert systems with applications*, **28**(4), 603–614.

PALEPU, KRISHNA G. 1986. Predicting takeover targets: A methodological and empirical analysis. *Journal of accounting and economics*, **8**(1), 3–35.

PEAT, MAURICE, & STEVENSON, MAXWELL. 2008. Predicting Australian takeover targets: a logit analysis. *In: The 6th INFINITI Conference on International Finance*.

PUSKÁS, CSABA, & DANCS, ISTVÁN. 2002. *Vektorterek*. Aula Kiadó Kft.

REGE, UDAYAN P. 1984. Accounting Ratios to Locate Take-Over Targets. *Journal of Business Finance & Accounting*, **11**(3), 301–311.

RUBIN, DONALD B. 1987. Comment. *Journal of the American Statistical Association*, **82**(398), 543–546.

STEVENS, DONALD L. 1973. Financial characteristics of merged firms: A multivariate analysis. *Journal of Financial and Quantitative Analysis*, **8**(2), 149–158.

TIKHOMIROV, VLADIMIR M. 1996. The evolution of methods of convex optimization. *The American Mathematical Monthly*, **103**(1), 65–71.

VAPNIK, VLADIMIR. 2013. *The nature of statistical learning theory*. Springer science & business media.

ZMIJEWSKI, MARK E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting research*, 59–82.