

A modellszelekció kérdései

Ferenci Tamás

tamas.ferenci@medstat.hu

Utoljára frissítve: 2023. május 9.

Tartalom

Tartalomjegyzék

1	Általánosítóképesség, túlilleszkedés	1
2	Modellszelekció	11
2.1	A modellszelekció tartalma	11
2.2	Modellszelekciós tesztek	11
2.2.1	Kitérő: modellezési filozófiák	12
2.3	Modellszelekciós mutatók, kritériumok	13

1. Általánosítóképesség, túlilleszkedés

Pár gondolat a magyarázó változók körének kiválasztásához

- Eddig egyetlen minősítőt láttuk egy modell jóságának: az R^2 -et
- Tételmondat: új változó bevonásával R^2 értéke *mindenképp* nő (de legalábbis nem csökken), teljesen függetlenül attól, hogy mi a bevont változónk, mik vannak már a modellben stb.
- Tehát: ha az R^2 -tel jellemezzük a modellünket, akkor *mindig* az összes potenciális magyarázó változó felhasználása lesz a legjobb döntés
- A valóságban azonban már nem biztos!
- Mert: az R^2 a *minta* jó leírását jellemzi, de mi a sokaságot akarjuk megragadni
- A kettő ellentmondásba kerülhet!

A tételmondat indoklásaként gondoljunk arra, hogy „legeslegrosszabb esetben” az újonnan bevont változó együttthatójára nulla mindenképp becsülhető – ekkor pedig ESS szempontjából pont ott vagyunk, mint az eredeti modell esetében!

Általánosítóképesség

- Azt, hogy a modell – a mintából kinyert információk alapján – mennyire jól tud a sokaságról (tehát a mintán kívüli világról) is számot adni, *általánosítóképességnek* nevezzük
- Igazából mi erre játszunk!
- ... ennyiben (erre a célra) az R^2 nem szerencsés mutató

Az R^2 a minta jó „megjegyzését” mutatja. Ez nekünk nem öncél – gondoljunk bele: ha csak a mintát akarnánk megjegyezni, akkor kár is regressziós modellt alkotni, használhatnánk egyszerűen magát a mintát is, ami ugye a rendelkezésünkre áll...

Általánosítóképesség

- Persze az sem jó megközelítés, hogy az R^2 -tel nem törődünk, hiszen ha nem szedünk ki elég információt a mintából, akkor sem várható, hogy a sokaságról jól tudunk nyilatkozni (mivel arra vonatkozóan csak a mintára támaszkodhatunk)
- Tehát: kompromisszumra van szükség a mintainformációk felhasználásában...
 - ... ha túl keveset használunk fel, akkor nem nyerünk elég jó képet a sokaságról
 - ... ha túl sokat használunk fel, akkor túlságosan „ráfókuszálunk” a mintára

Általánosítóképesség

- Ahogy egyre több információt nyerünk ki a mintából (egyre jobban „elköteleződünk” mellette), úgy egy pontig javul, majd ezen túl romlik az általánosítóképesség
- Tehát: nem csak nem javít a több információ, de egyenesen ront (ezért az „ellentmondás”)!

Alulilleszkedés, túlilleszkedés

- A fentiek jól értelemezhetőek a *gépi tanulás* fogalomkészletével
- Itt a tanulás információkinyerés a mintából
- Ha ezt túl kis mértékben hajtjuk végre, akkor alulilleszkedésről (alultanulásról)...
- ... ha túl nagy mértékben, akkor túlilleszkedésről (túltanulásról) beszélünk
- A túltanított modell látszólag nagyon jó (a mintát jól megragadja), de valójában nem az, mert a mintán kívüli képességei gyatrák lesznek (hiszen túlságosan „ráfókuszált” a mintára)

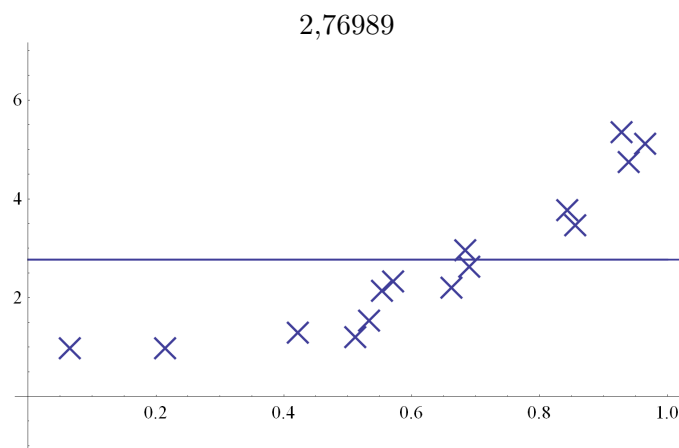
Egy példa a túlilleszkedésre

- Egyszerű kétváltozós feladat: egy magyarázó- és egy eredményváltozó
- A példánkban a tanítás fokát tehát nem a magyarázó változók számával fogjuk mérni, hanem a függvényforma bonyolultságával: $Y = \beta_1 + \beta_2 X + u$, $Y = \beta_1 + \beta_2 X + \beta'_2 X^2 + u'$, $Y = \beta_1 + \beta_2 X + \beta'_2 X^2 + \beta''_2 X^3 + u''$ stb.
- Tehát az eredményváltozót a magyarázó változó egyre nagyobb fokszámú polinomjával közelítjük (a polinom fokszámát jelölje p)
- (A függvényforma ilyen megválasztásával később foglalkozunk részleteiben, de most nem is ez a lényeg)

Egy példa a túlilleszkedésre

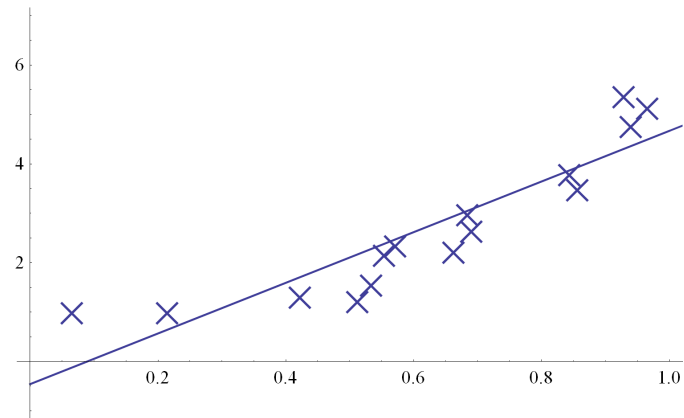
- Hogy tudjuk mi a „jól illeszkedő” modell, elárulom, hogy az adatokat valójában egy $Y = 5 \cdot X^3 + 1 + u$ modell szerint generáltam, ahol $u \sim \mathcal{N}(0; 0,3)$
- Tehát lényegében: „zajos harmadfokú” függvény
- A jól illeszkedő modell – ezt *most* tudjuk, általában persze nem! – a harmadfokú lenne

Alulilleszkedés: $p = 0$



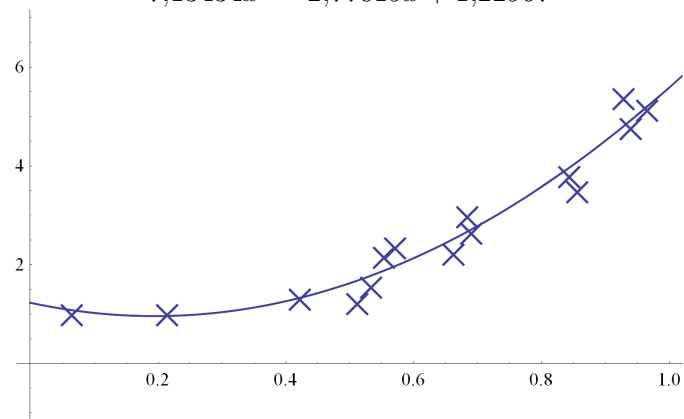
Alulilleszkedés: $p = 1$

$$5,12654x - 0,458165$$



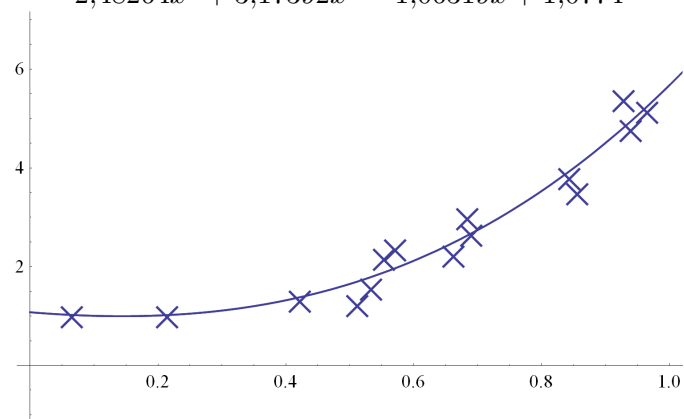
Nagyjából jó illeszkedés: $p = 2$

$$7,13434x^2 - 2,77819x + 1,22967$$

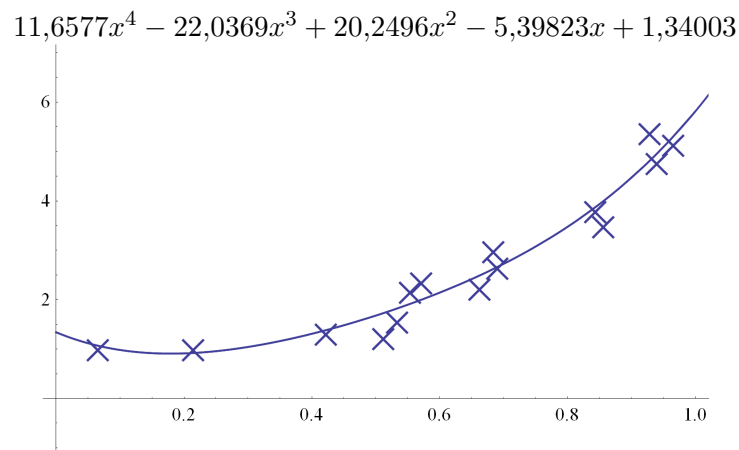


Nagyjából jó illeszkedés: $p = 3$

$$2,48264x^3 + 3,17392x^2 - 1,06319x + 1,0774$$

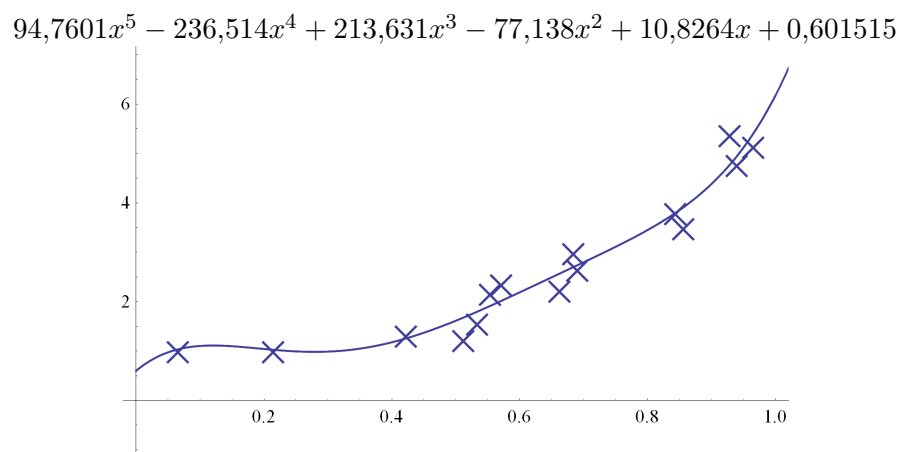


Nagyjából jó illeszkedés: $p = 4$



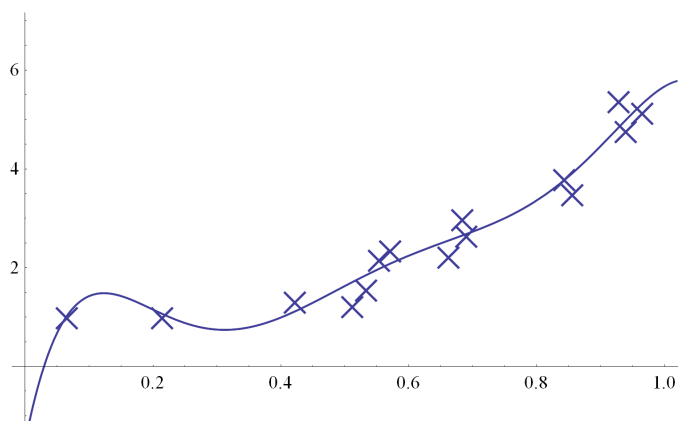
Érdekes észrevenni, hogy még ha tekintettel vagyunk az általánosítóképességre, akkor sem tudjuk mintából egyértelműen megmondani, hogy pont a $p = 3$ a jó válasz. Ez azonban nem meglepő, sőt, épp várható: egymáshoz közeli lehetőségek minta alapján nem feltétlenül különíthetők el, ha nem elég nagy a mintanagyság. Minél kisebb a differencia, annál nagyobb mintaméret kell, hogy megbízhatóan el tudjuk dönteni, hogy a szóba jövő lehetőségek közül melyik a valódi modell – ez nyilván általános statisztikai elv. Kevés pont alapján – sajnos – elképzelhető, hogy nem eldönthető, hogy lineáris, négyzetes, vagy épp exponenciális görbe a valódi modell.

Túlilleszkedés: $p = 5$



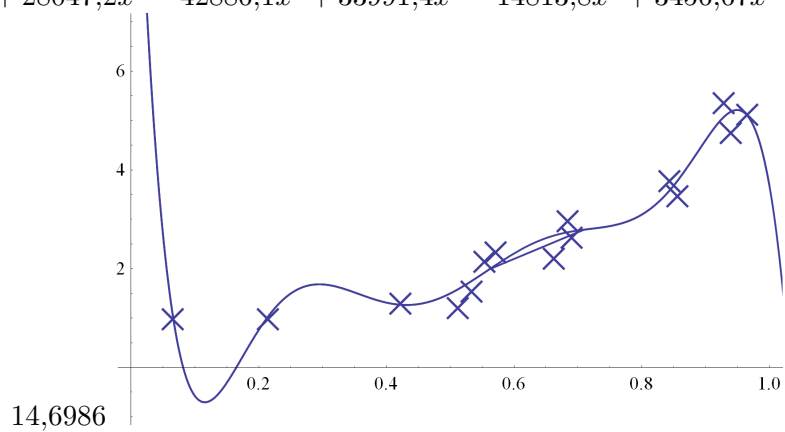
Túlilleszkedés: $p = 6$

$$-556,426x^6 + 1895,28x^5 - 2494,87x^4 + 1587,69x^3 - 489,325x^2 + 64,8299x - 1,52203$$



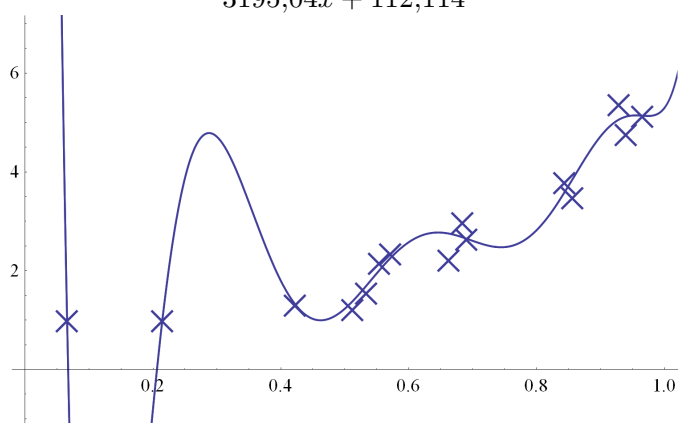
Túlilleszkedés: $p = 7$

$$-7426,18x^7 + 28047,2x^6 - 42886,1x^5 + 33991,4x^4 - 14813,8x^3 + 3456,67x^2 - 380,286x +$$



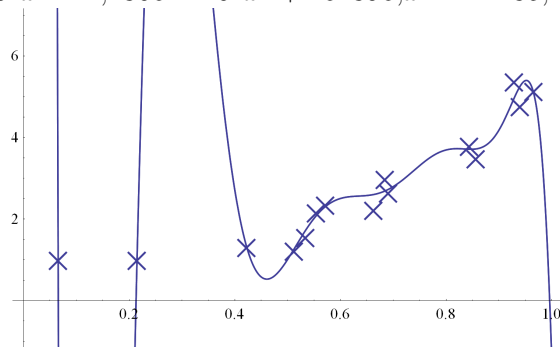
Túlilleszkedés: $p = 8$

$$59039,2x^8 - 282296,x^7 + 565254,x^6 - 613881,x^5 + 390937,x^4 - 146967,x^3 + 31001,6x^2 - 3195,04x + 112,114$$



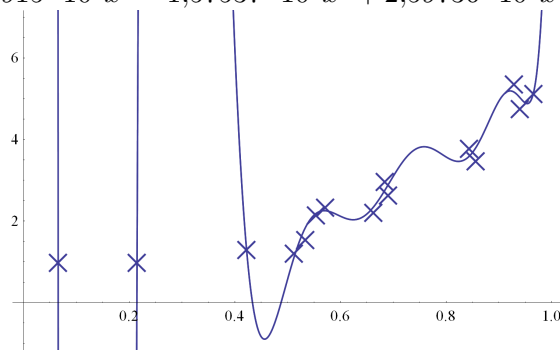
Túlilleszkedés: $p = 9$

$$-722495, x^9 + 3,85053 \cdot 10^6 x^8 - 8,84295 \cdot 10^6 x^7 + 1,1426 \cdot 10^7 x^6 - 9,08926 \cdot 10^6 x^5 + 4,57009 \cdot 10^6 x^4 - 1,43064 \cdot 10^6 x^3 + 262396, x^2 - 24485,1 x + 807,137$$



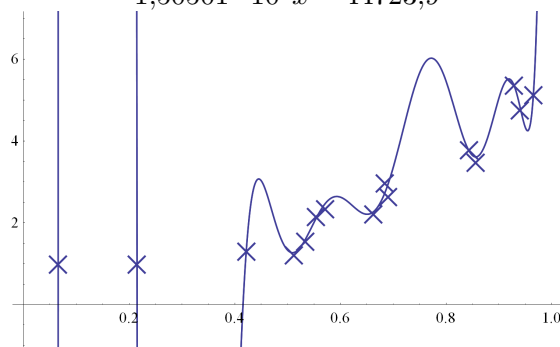
Túlilleszkedés: $p = 10$

$$8,61299 \cdot 10^6 x^{10} - 5,24999 \cdot 10^7 x^9 + 1,40371 \cdot 10^8 x^8 - 2,16006 \cdot 10^8 x^7 + 2,1085 \cdot 10^8 x^6 - 1,35546 \cdot 10^8 x^5 + 5,75915 \cdot 10^7 x^4 - 1,57537 \cdot 10^7 x^3 + 2,59736 \cdot 10^6 x^2 - 223991, x + 7044,46$$



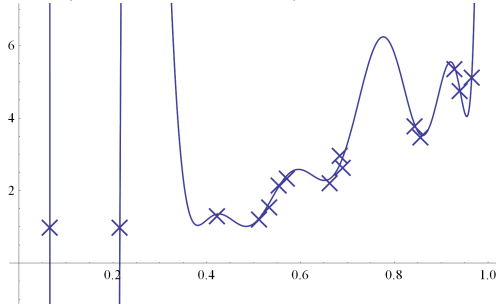
Túlilleszkedés: $p = 11$

$$9,81027 \cdot 10^7 x^{11} - 6,54761 \cdot 10^8 x^{10} + 1,94347 \cdot 10^9 x^9 - 3,37777 \cdot 10^9 x^8 + 3,80722 \cdot 10^9 x^7 - 2,91 \cdot 10^9 x^6 + 1,53045 \cdot 10^9 x^5 - 5,49469 \cdot 10^8 x^4 + 1,30416 \cdot 10^8 x^3 - 1,91189 \cdot 10^7 x^2 + 1,50501 \cdot 10^6 x - 44723,9$$



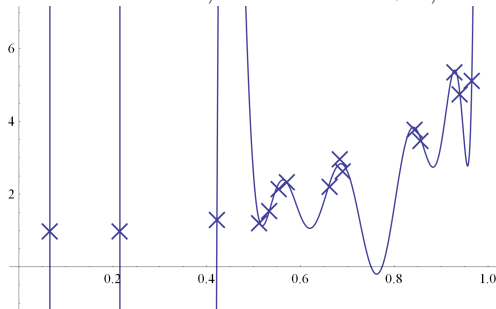
Túlilleszkedés: $p = 12$

$$1,97286 \cdot 10^8 x^{12} - 1,37728 \cdot 10^9 x^{11} + 4,31319 \cdot 10^9 x^{10} - 7,99714 \cdot 10^9 x^9 + 9,75531 \cdot 10^9 x^8 - 8,22533 \cdot 10^9 x^7 + 4,8983 \cdot 10^9 x^6 - 2,06632 \cdot 10^9 x^5 + 6,08915 \cdot 10^8 x^4 - 1,211 \cdot 10^8 x^3 + 1,51977 \cdot 10^7 x^2 - 1,05188 \cdot 10^6 x + 28665$$



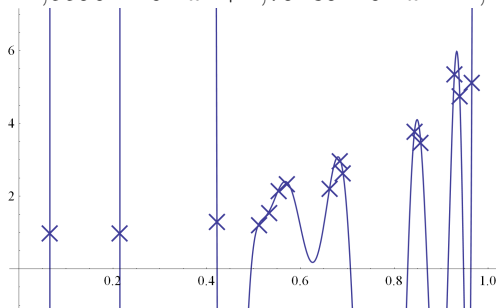
Túlilleszkedés: $p = 13$

$$1,33188 \cdot 10^{10} x^{13} - 1,09101 \cdot 10^{11} x^{12} + 4,06208 \cdot 10^{11} x^{11} - 9,08859 \cdot 10^{11} x^{10} + 1,36095 \cdot 10^{12} x^9 - 1,43708 \cdot 10^{12} x^8 + 1,0978 \cdot 10^{12} x^7 - 6,12006 \cdot 10^{11} x^6 + 2,4775 \cdot 10^{11} x^5 - 7,14241 \cdot 10^{10} x^4 + 1,41049 \cdot 10^{10} x^3 - 1,77685 \cdot 10^9 x^2 + 1,24223 \cdot 10^8 x - 3,41822 \cdot 10^6$$

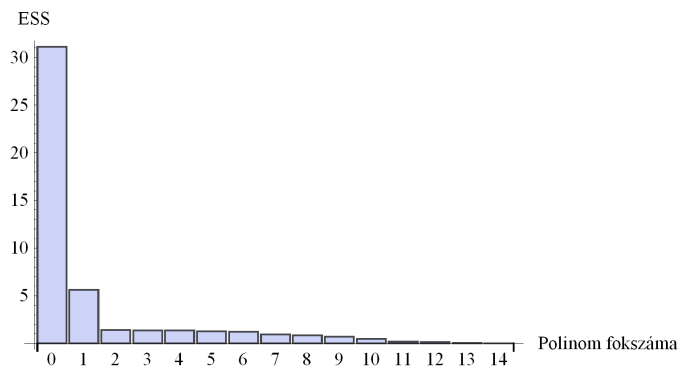


Túlilleszkedés: $p = 14$

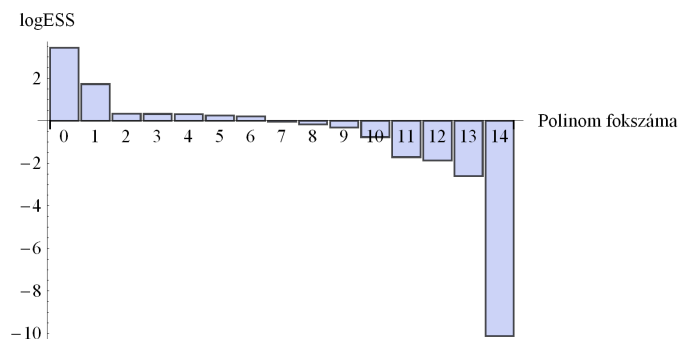
$$2,23808 \cdot 10^{11} x^{14} - 1,95447 \cdot 10^{12} x^{13} + 7,81606 \cdot 10^{12} x^{12} - 1,89512 \cdot 10^{13} x^{11} + 3,10833 \cdot 10^{13} x^{10} - 3,64245 \cdot 10^{13} x^9 + 3,1386 \cdot 10^{13} x^8 - 2,01508 \cdot 10^{13} x^7 + 9,65479 \cdot 10^{12} x^6 - 3,41996 \cdot 10^{12} x^5 + 8,76076 \cdot 10^{11} x^4 - 1,55904 \cdot 10^{11} x^3 + 1,79489 \cdot 10^{10} x^2 - 1,16536 \cdot 10^9 x + 3,04682 \cdot 10^7$$



Hiba az egyes fokszámok mellett



Jobban láthatóan...



Itt a függőleges tengely logaritmikus beosztású, hogy a nagyon kis számok tartományában is látszódnának a változások.

A túlilleszkedés hatása

- Itt a tanítás mértékét a polinom fokszáma jelzi
- A példa tökéletesen mutatja, hogy mi a túlilleszkedés tartalma:
 - A mintaadatokat ugyan egyre jobban megtanuljuk...
 - ... de közben a mintán kívüli világról egyre kevesebbet tudunk mondani (holott minket ez érdekelne igazából!)
- A túltanulás igazi problematikáját az adja, hogy ez utóbbi *elkerülhetetlenül* bekövetkezik, ha a tanítást túl sokáig folytatjuk (az ellentmondás a két szempont között, ugyebár)

Túlilleszkedés túl sok magyarázó változó miatt

- A magyarázó változók száma tipikus példája a tanítás fokának
- Túl kis mértékű tanítás (túl kevés magyarázó változó) esetén az alulilleszkedés miatt lesz rossz a modellünk...
- ... túl nagy mértékű tanítás (túl sok magyarázó változó) esetén a túlilleszkedés, az általánosítóképesség leromlása miatt
- Szemléletes megjelenés: a bevont magyarázó változók száma csökkenti a tesztek szabadsági fokainak számát (erre ugyanis sokszor jön elő valamilyen $n - (k + 1)$ jellegű kifejezés), így az erejüket; „leköti a szabadsági fokokat”
- Az R^2 ezt nem jellemzi, csak a mintához való illeszkedést
- Valahogy „javítani” kell; ezzel fogunk most foglalkozni

Megoldási lehetőségek I.

- Magyarázó változók számának csökkentése *csak* a bennük lévő információk alapján, tehát *nem* is nézve az eredményváltozót („blinded to the outcome”)
 - A legtisztább megoldás
 - Két alapvető kivitelezési lehetőség: szakmai szempontok szerinti szűrés, vagy statisztikai alapú redundanciavizsgálat a magyarázó változók körében és redundánsak elhagyása vagy összevonása
 - Ebben segíthetnek az arra vonatkozó irányelvek, hogy adott mintanagyság mellett mennyi prediktor modellezhető
- Minden magyarázó változó felhasználása, de a regresszió regularizálása (penalizálás)
- Egyéb korszerű megoldások (pl. bayes-i modellátlagolás, BMA)

Megoldási lehetőségek II.

- Statisztikai alapú szűrés
 - Ezzel fogunk most részletesen foglalkozni
 - De vigyázat, ész nélkül nem használható, mert az *maga* is túlilleszkedéshez vezethet!
 - Ész nélkül: össze-vissza mindenféle lehetőséget megvizsgálva, hogy melyik jobb; ehelyett vezessenek minket amennyire lehet szakmai megfontolások, a próbálkozások lehetőleg legyenek pre-specifikáltak (ne az adatok sugallják őket), és ha kétség van, inkább közöljünk többféle modellt

2. Modellszelekció

2.1. A modellszelekció tartalma

A modellszelekció fogalma

- Modellszelekció alatt az optimális magyarázó változó-kör meghatározását értjük
- Ennek megfelelően foglalkozik változó bevonásának/elhagyásának hatásával...
- ...de nem „mikroszkopikusan” (mi történik a többi változó becsült paramétereivel stb.), hanem „makroszkopikusan” (mi történik a modell jóságával)
- Az előbbi inkább a modellspecifikáció kérdése, később fogunk vele foglalkozni
- Továbbá: a modellspecifikációhoz szoktuk sorolni az adott magyarázó változó-kör melletti függvényforma kialakítást (de nincs egyértelmű határ a kettő között)

A modellszelekció problematikájának megoldása

- Az biztos, hogy a mintához való illeszkedés az R^2 -tel jellemezhető
- Innentől két lépésben lehet továbbhaladni a modellszelekcióval:
 1. Két modell között úgy döntünk, hogy megnézzük, hogy lényeges-e köztük az R^2 -beli különbség... és csak akkor választjuk a bővebbet, ha nem csak nagyobb (ez biztos), de *lényegesen* nagyobb az R^2 -e (más szóval: egy modellből mindazon változókat elhagyjuk, melyek nem csökkentik *lényegesen* az R^2 -et, még ha számszerűen csökkentik is)
 2. Definiálunk olyan mutatót az R^2 helyett, mely az R^2 -hez hasonlóan figyelembe veszi a mintához való illeszkedést, de – azzal szemben – az ehhez szükséges magyarázó változók számát *is*
- Most e két kérdést fogjuk közelebbről is megvizsgálni

Itt (és mindenhol máshol is) a „lényeges”-et a „statisztikailag szignifikáns” szinonimájaként használjuk, tehát természetesen úgy értjük, hogy mintavételi értelemben lényeges, azaz olyan mértékű, ami nem egyeztethető össze a mintavételi ingadozással: adott szignifikanciaszinten nem hihető, hogy a változás pusztán a mintavételi ingadozásnak tudható be, ezzel szemben feltehető, hogy tényleges sokasági különbség van a háttérben.

2.2. Modellszelekciós tesztek

A modellszűkítésről

- Már láttuk, hogy miért akarhatunk modellt szűkíteni (változót elhagyni a modellből), még ha ezzel rontunk is az R^2 -en (és még látni fogunk más okot is)

- Melyik változót lehet érdemes ezek miatt elhagyni? → *mérlegelés* a fentiekben javulás és az R^2 romlása között
- Sok vagy kevés a romlás? – a szó statisztikai értelmében lényeges-e!
- Azaz túlmutat-e a mintavételi ingadozáson: ehhez teszt kell

Nested (beágyazott) modellszelekció: a szűkebb modell minden változója benne van a bővebb modellben – elhagyhatóak anélkül, hogy a modell lényegesen romlana (azt is jelenti, hogy a két R^2 között nincs lényeges különbség)

2.2.1. Kitérő: modellezési filozófiák

Az LM és a Wald-teszt eltérései

- Ha ugyanazt a hipotézist vizsgálják, mi a különbség köztük?
- A nyilvánvaló: teljesen más elven épülnek fel
- Ennek konkrétabb következményei:
 1. Nem feltétlenül ugyanakkor utasítanak el; sőt, ennél több is mondható: az LM-próba *mindig* az elfogadás felé „hajlik” (olyan értelemben, hogy ha ez elutasít, akkor a Wald is, viszont ha a Wald elfogad, akkor az LM is elfogad)
 2. A Wald kismintás próba, az LM-próba nagymintás (értsd: tulajdonságai csak aszimptotikus értelemben garantáltak), de azért a gyakorlatban már néhány-szor 10 mintaelemre is elég jól szokott közelíteni
 3. Belátható, hogy a Wald-teszt csak a korlátozatlan, az LM-teszt csak a korlátozott modell becslését igényli; ez utóbbi egyszerűbb (gyakorlatban számít!)

A „kismintás” természetesen nem azt jelenti, hogy a teszt csak kis mintán működik, hanem épp ellenkezőleg: azt, hogy *minden* mintanagyság mellett működik (szemben a nagymintás teszttel, ami *csak* nagy mintán működik!). Talán szerencsésebb is ezért a „véges mintás” kifejezés, mely egyúttal a különbségtétel valódi okára is utal: a kismintás tesztek tulajdonságai „ $\forall n$ ”, míg a nagymintások „ $\lim_{n \rightarrow \infty}$ ” értelemben garantáltak.

Az LM és a Wald-teszt eltérései

- Van egy általánosabb különbség is: más modellezési filozófiához illeszkednek
- A Wald-teszt inkább az „általánostól az egyszerűig” filozófiának (Hendry/LSE) felel meg (a korlátozatlan modellből indul, és kérdezi, hogy lépünk-e a csökkentés irányába)
- Az LM-próba inkább az „egyszerűtől az általánosig” filozófiának felel meg (a korlátozott modellből indul, és kérdezi, hogy lépünk-e a bővítés irányába)

- ... hát ez a különbség – hiába ugyanaz *formailag* a hipotézispár!

Nem igazán lehet válaszolni arra a kérdésre, hogy melyik a „jobb” modellezési filozófia: nagyon sok, részben egymásnak ellentmondó, elméleti és gyakorlati szempont merül fel a választásnál. Ezzel a kérdéssel könyvtárnyi irodalom foglalkozik.

Az LM és a Wald-teszt eltérései

- Már most megjegyezzük, hogy az „újonnan felvett” változó nem szükségszerű, hogy még nem szereplő változó legyen: lehet egy már bent levő változó valamilyen új, nemlineáris függvényformája (pl. négyzete), vagy változók interakciója (ld. később)
- E célra általában LM-tesztet használnak, emiatt igaz az, hogy az LM-elvű teszteket kicsit általánosabban is használják az ökonometriában, más hipotézisek tesztelésére is
- ... tehát: ez modellspecifikációs tesztként is felhasználható!

2.3. Modellszelekciós mutatók, kritériumok

Az R^2 „megjavítása”

- Ahogy láttuk az R^2 *önmagában* nem minősít egy modellt, mert csak a hibát minimalizálja, a túl sok változó káros hatásával egyáltalán nem foglalkozik („egyoldalú” mérlegelés)
- Nem lehetne ezt valahogy kijavítani? → olyan mutatót konstruálni, ami mindkét szempontra tekintettel van
- Ötlet: induljunk ki az R^2 -ből, de büntessük a magyarázó változók számának növelését
- Bár máshonnan származik, de épp ennek a logikának felel meg (gondoljuk végig!) a *korrigált* R^2 :

$$\bar{R}^2 = 1 - \left(1 - R^2\right) \frac{n - 1}{n - k - 1}$$

- Ez már alkalmas különböző számú magyarázó változót tartalmazó modellek összehasonlítására

A korrigált R^2 klasszikus bevezetése szerint $1 - \bar{R}^2$ megegyezik a (sokasági) hibatag és az eredményváltozó becült varianciáinak a hányadosával. A „sima” R^2 -et ugyanis úgy definiáltuk mint $R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$, ami helyett azt is írhattuk volna, hogy $R^2 = 1 - \frac{ESS/n}{TSS/n}$, ez utóbbiból még jobban látszik, hogy két szórásnégyzet hányadosától függ. A probléma, hogy belátható módon sem az ESS/n , sem a TSS/n nem torzítatlan becslője a megfelelő szórásnégyzetnek (ez utóbbi klasszikus induktív statisztikai felismerés) – az

$ESS/(n - k - 1)$ és a $TSS/(n - 1)$ viszont az. Ezeket behelyettesítve az előbbi képletbe nyerjük a \bar{R}^2 -et.

Ilyen szempontból – és ez is egy nagyon fontos megállapítás önmagában is! – a korrigált R^2 azért is érdekes, mert a szokásos R^2 következtető statisztikai értelemben torzított (gondoljuk végig, hogy nem is lehet más, hiszen 0 és 1 között van mindig – mi van, ha a valódi R^2 0?), a korrigált R^2 épp ezen a torzításon javít! (Sajnos nem oldja meg teljesen: igaz, hogy a nevezőre és a számlálóra is torzítatlan becslést használ, de egy hányados nem biztos, hogy torzítatlan attól, mert a nevezője és a számlálója is torzítatlan.)

Az \bar{R}^2 főbb tulajdonságai

- $\bar{R}^2 \leq R^2$
- Ebből következően 1-nél nem lehet több...
- ...de 0-nál lehet kisebb (ha sok magyarázó változóval is csak gyenge magyarázást (kis R^2 -et) tud elérni)
- Ez már csökkenhet is új változó bevonásával (belátható, hogy ez a változó t -hányadosától függ)
- Ilyen módon már nem beágyazott modellek szelekciójára is használható...
- ... de vigyázat: csak akkor, ha az eredményváltozó azért ugyanaz (különben a megmagyarázandó variancia is más lenne)

Automatikus modellszelekció

- Megadjuk a változók egy maximális halmazát (l darab potenciálisan szóba jövő magyarázó változó), és „a gép” kiválasztja, hogy melyik részhalmaza az optimális: melyeket érdemes egy modellbe bevonni, hogy az a legjobb legyen
- Jóság valamilyen célfüggvény szerint (ami ugye *nem* R^2 , hogy a dolognak értelme is legyen, hanem pl. \bar{R}^2)
- Léteznek heurisztikus stratégiák (mind mohó algoritmus), hogy ne kelljen a 2^l kombinációt tesztelni (forward, backward, stepwise szelekció)
- Az automatikus modellszelekció használata azonban szinte *minden esetben* és *határozottan ellenjavallt*, az alkalmazásával nyert modelleknek torzítottak lesznek a regressziós koeficiensei, torzítottak lesznek a becsült standard hibái, ebből adódóan torzítottak lesznek a konfidenciaintervallumai, a szokásos p -értékek falsak lesznek, a rájuk alapozott tesztek invalidak, a t és F statisztikáknak nem t illetve F eloszlásuk lesz, torzított lesz a modell R^2 -e stb.

Információs kritériumok

- Vannak további mutatók is, melyek egyszerre büntetik a magyarázó változók nagy számát és a nagy hibát, a kettő között egyensúlyt keresve, pl.
 - Akaike (AIC): $AIC = \frac{ESS}{n} e^{\frac{2(k+1)}{n}}$
 - Schwarz (SBC): $BIC = \frac{ESS}{n} n^{\frac{k+1}{n}}$
 - Hannan-Quinn (HQC): $HQC = \frac{ESS}{n} (\ln n)^{\frac{2(k+1)}{n}}$
- Teljesen más elven (információelméleti alapon) épülnek fel mint az \bar{R}^2
- Hiba jellegű mutatók, ezért őket *minimalizálni* akarjuk és nem maximalizálni!
- Sok van belőlük, döntsük el előre, hogy melyiket használjuk a modellszelekcíóra!
- Ezekkel nem csak beágyazott modellek hasonlíthatóak össze (de azért jobbak a tulajdonságaik ilyenkor)

Modellszelekciós stratégiák

- Itt már látszik, hogy miért mondtuk az elején, hogy az ökonometriai munka iteratív
- Diagnosztizáljuk a modellt, és – ha ilyen baj van vele – szűkítjük vagy bővítjük, majd újra diagnosztizáljuk, majd...
- De vigyázat: újra fontos felhívni a figyelmet, hogy ha rengeteg ilyen iterációra kerül sor az értelemszerűen *maga is* túlilleszkedéshez vezethet (túlságosan rá szabjuk a konkrét mintára a modellt)!