

POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

Název: Řešení soustavy lineárních rovnic v nestandardních podmínkách

Autor: Šimon Vojtek

SHRNUTÍ OBSAHU PRÁCE

Práce je věnována metodám řešení soustav lineárních rovnic, s nimiž se ve statistice setkáváme v kontextu odhadu parametrů regresních modelů. Problematika je představena v první kapitole nazvané netradičně jako „Předmluva“. Druhá kapitola je většinou věnována metodě nejmenších čtverců pro odhad parametrů lineárního modelu, pojednáno je též o nelineární regresi a stručně též o L_1 odhadu v lineárním modelu. Třetí kapitola se z větší části snaží o shrnutí metod používaných k výběru proměnných lineárního modelu. Práce je zakončena závěrem. Zde v jednom jediném odstavci zvládl autor (kromě jiných pravopisně-stylistických prohřešků) zvolit dvakrát chybně i/y ...

CELKOVÉ HODNOCENÍ PRÁCE

Téma práce. Téma hezkým způsobem propojuje metody maticové lineární algebry a numerické matematiky, jež jsou většinou vyučovány v povinných kurzech programu *Obecná matematika* s regresními modely ve statistice, s nimiž se student bakalářského stupně dosud setkal pouze okrajově. Téma proto poskytovalo potenciál shrnout koherentním způsobem studovanou problematiku na úrovni přístupné studentovi programu *Obecná matematika*, zaměření *Stochastika* s tím, že si student musel některé pasáže samostatně dostudovat. Jednalo se tudíž o ideální téma bakalářské práce. Autor bohužel zmíněný potenciál využil pouze částečně, cíle práce bylo do jisté míry dosaženo, mám nicméně značné výhrady ke kvalitě zpracování, viz níže.

Vlastní příspěvek. Vlastní příspěvek autora spočívá do jisté míry v kompilaci metod z různých oblastí s potenciálem tyto shrnout ve formě koherentního a srozumitelného textu. Pouze částečně (spíše vůbec) se podařilo vytvořit text, který by byl srozumitelný pro (i lepšího) studenta 3. ročníku bakalářského programu *Obecná matematika*. Konstatuji, že bez poměrně pokročilé znalosti dané problematiky bych mnohdy ani já netušil, co chce autor danou pasáží čtenáři sdělit. Jako příklad uveďme první odstavec sekce 2.3, který měl patrně zavést model/problematiku lineární regrese:

Jde o metodu, která aproximuje body křivkou viz (Yan a Su, 2009), a lze ji vyjádřit jako lineární kombinaci n zadaných funkcí. Tedy hledanou neznámou funkci budeme hledat různými metodami, v konkrétním určitém tvaru $\hat{y} = f(z, \xi_1, \xi_2, \dots, \xi_n) = \xi_1 f_1(z) + \xi_2 f_2(z) + \dots + \xi_n f_n(z)$, kde reálný vektor neznámých ve tvaru $[\xi_1, \xi_2, \dots, \xi_n] \in \mathbb{R}^n$ a f_1, f_2, \dots, f_n jsou zadané funkce z -tu (často se za f_1 bere konstantní funkce) a \hat{y} vyjadřuje predikovanou hodnotu pro dané z .

Věřím nicméně, že autor si jednotlivé poznatky alespoň pro sebe utřídil. Souvislý text stvořen byl a lze ho prohlásit za vlastní příspěvek autora.

Matematická úroveň. Matematická úroveň textu je na mnoha místech bídná. Ne vždy je jasné, z čeho autor vychází, co předpokládá a co tvrdí. Vybírám z matematických nedostatků.

1. Za nejzásadnější problém považuji fakt, že autor téměř nerozlišuje situace, kdy se zabývá čistě řešením soustav rovnic (až do konce sekce 2.2), kde není žádná náhodnost a kdy (navíc) řeší problematiku odhadu parametrů a též statistických vlastností těchto odhadů (od sekce 2.3), kde je zapotřebí uvést též předpoklady o stochastickém chování jednotlivých veličin. V první řadě

je nutné sdělit čtenáři, které veličiny jsou náhodné a které nikoliv. Tento problém se projevuje hned od začátku sekce 2.2, kde se autor začíná věnovat lineární regresi. Pouze si domýšlím (s využitím znalosti problematiky), jaká (náhodná) veličina/vektor reprezentuje data a jaký přesně stochastický (statistický) model pro příslušná data se předpokládá. Není mi nicméně jasné, proč se ke konci strany 8 podmiňuje maticí A . Je náhodná? Předpokládá se o její náhodnosti něco? A pokud je náhodná, nemá náhodou $\text{var}(x_0)$ ve výrazu (2.3) na str. 9 být též podmíněn touto maticí?

2. Nejasnost v tom, co jsou data reprezentovaná náhodným vektorem, co je předpokládaný model a co jsou neznámé parametry se projevuje též např. na str. 18 v popisu „Algoritmu“. Krok 1 zní: „Najít \mathcal{M}_0 nulový model, který nemá žádný parametr bude odhadovat pouze výběrový průměr.“ Když model nemá žádný parametr, co se tedy odhaduje (nehledě na to, že model samotný nikdy nic neodhadne)? A je potřeba odhadovat výběrový průměr?
3. Definice 6 na str. 9 je podivnou kombinací definice a tvrzení. Autor patrně zamýšlel definovat pojem nejlepšího nestranného odhadu. Současně však vyslovil tvrzení o tom, že dříve zavedený odhad x_0 je nejlepším nestranným odhadem vektoru x_x neznámých parametrů. Toto tvrzení (Gaussova-Markovova věta) je následně dokazováno.
4. V sekci 2.4 pojednává autor o hledání odhadu metodou nejmenších čtverců v případě nelineárního modelu. Autor se dále snaží vysvětlit fungování Newtonovy metody pro nalezení příslušného minima. Text na str. 12 v podstatě sleduje správný postup, ale jeho zahrnutí v dané podobě do bakalářské práce na matematickém studijním programu ve mně vyvolává zoufalství. Autor kupí myšlenky a prokládá je zjištěními „... takže musíme předpokládat, že je regulární.“, „... tak musíme předpokládat, že r_i jsou dost malé a druhé parciální derivace nejsou velké, což si můžeme vyjádřit...“ Ocenil bych, pokud by příslušná zjištění byla shrnuta na jednom místě ve formě předpokladů, klidně i na konci dané sekce po předvedených úvahách.
5. Možná jde o detail, ale... V důkaze Tvrzení 1 (str. 4) se používá klasický symbol pro jednoduchou derivaci (apostrof/čárka). Je si autor vědom faktu, že v tomto případě je zapotřebí uvážit parciální derivace vzhledem k n složkám vektoru x a typicky $n > 1$?
6. Úsměvný je výraz „výchylka“ používaný v práci pro (náhodou) odchylku.

Práce se zdroji. Zdroje, z nichž práce vychází, jsou citovány, ale... Seznam použité literatury na str. 25 trpí mnoha formálními nedostatky a nekonzistencemi, celkově se jedná o nedodělek. Začátky slov v názvech anglicky psaných knih jsou tu uvedeny s velkými písmeny (jak je obvyklé), tu s malými, u některých chybí místo vydání, u některých je uvedeno ISBN, u některých nikoliv. V případě odkazu na práci C. F. Bauma není vůbec zřejmé, že se jedná o učební text ke kurzu na jisté univerzitě. U internetových zdrojů není uveden datum přístupu. Není mi též zřejmé, co znamená „[u.a.], [2. print.]“ u odkazu Birkes a Dodge (1993).

Formální úprava. Formální úprava práce je nedobrá. V prakticky celé práci se na téměř každé straně vyskytuje několik z typově následujících nedostatků.

1. Náhodné střídání množného a jednotného čísla v obratech typu uděláme/udělám, např. „... kde si nejdříve ukážeme, že je lze ekvivalentně přeformulovat, což nám usnadní vizualizaci problému a to předvedu na vytvořených obrázcích.“ na str. 2 (a potom i dále).
2. Ne zcela spisovná vyjádření – hned v poděkování vedoucímu práce na str. ii autor děkuje za „dávání cenných rad“. Na str. 13 mě mírně děsí věta „Uděláme substituci za ... a dostaneme funkci ..., která dělá lineární kombinaci funkcí ...“. Některá vyjádření jsou až úsměvná, např. „Proto musíme zvolit způsob, jako například (minimalizace ℓ_2 -normy) pro získání jednoho řešení, o kterém ani nevíme, zdali je způsob zvolený správně.“ na str. 17.
3. Pravopisné chyby – interpunkční znaménka („V této bakalářské práci se budeme zabývat, hledáním ideálního řešení soustav lineárních rovnic.“ na str. 3), tečky za číslovkami („Příklad 2.

(konzistentní)“ na str. 3), i/y („V jednom z mých příspěvků...“ nebo „pomocí Newtonovi metody“ na str. 24) atp.

4. Jednopísmenné předložky, resp. spojka „a“ na konci řádku (6× hned na str. 2 a potom i dále).
5. Použití zkratk bez jejich zavedení („MNČ“ na str. 2).
6. Podivný znak pro uvozovky (např. na str. 3 na druhém řádku odspodu a dále).

S ohledem na poměrně chabou stylistickou a pravopisnou úroveň mohu nicméně alespoň konstatovat, že předložená práce nebyla s vysokou mírou pravděpodobnosti vytvořena s pomocí prostředků umělé inteligence a jedná se o dílo živého člověka.

OTÁZKY/TÉMATA K DISKUZI

1. V Tvzení 1 (str. 4) se předpokládá matice A s **nezávislými sloupci**. O jaký typ nezávislosti se zde jedná (stochastický?). V jakém kroku důkazu tento předpoklad využíváte?
2. Vysvětlete, co přesně myslíte vyjádřením „V jednom z mých příspěvků bylo úplné podrobné rozepsání a doplněním definic v důkazu, že MNČ generuje nejlepší nestranný lineární odhad.“, které uvádíte v Závěru na str. 24.
3. Vysvětlete, co přesně myslíte vyjádřením „Ukázali jsme, že MNČ minimalizuje ℓ_2 -normu, kde jsem přispěl dokázání, že x_{MN} je kolmé na množinu řešení.“, které uvádíte v Závěru na str. 24.

ZÁVĚR

Jedná se o slabou práci, kterou nicméně ještě **doporučuji** uznat jako bakalářskou pro studijní program *Obecná matematika* na MFF UK.

doc. RNDr. Arnošt Komárek, Ph.D.

Katedra pravděpodobnosti a matematické statistiky
Matematicko-fyzikální fakulta Univerzity Karlovy

V Jindřichově Hradci dne 22. srpna 2024