

Oponentní posudek bakalářské práce:

SAMUEL ŘÍHA:

MAXIMALIZACE GINIHO KOEFICIENTU V BINÁRNÍ LOGISTICKÉ REGRESI

Uchazeč se v předložené práci zabývá logistickou regresí. Hlavním cílem je odvození odhadů založených na maximalizaci Giniho koeficientů a jejich porovnání s jinými metodami.

Práce obsahuje větší množství podrobně odvozených odvození, ve kterých jsem nenalezl žádné chyby. Zároveň však práce bohužel není příliš přehledná. Hlavní nedostatky jsou zejména:

1. Na straně 2 není jasné, jak se liší výrazy $q(x)$ a $\eta(x)$. Nejdříve se píše, že $q(x) = P(Y = 1)$, ale jen o 3 řádky dál stojí doslova, že to tak „přímo není“ a $q(x)$ má být odhadem $\eta(x) = P(Y = 1) \dots$
2. V poznámce na straně 5 není jasné, jak lze získat libovolně malou chybu odhadu.
3. V důkazu věty 2.1 se mluví o množině všech distribučních funkcí, ale vzápětí se používá distribuční funkce alternativního rozdělení. Opět zde není jasný rozdíl mezi výrazy $q(F)$, η a $q(\hat{F})$.
4. Ve větě 2.2 by mělo být jasně řečeno, co se myslí slovem “hladká” (pak by to nemuselo být v uvozovkách).
5. Str. 11, 3. řádek: jakou funkci zobecňuje mocninná ztrátová funkce?
6. Důkaz věty 2.3: uvedená podmínka, viz předposlední vzorec, je zřejmě splněná i pro $k = 0$.
7. V důkazu věty 2.4 se používá spojitost ω , která se nikde nepředpokládá.
8. Na str. 23 se předpokládá normální rozdělení skóre. Nemělo by být různé rozdělení skóre pro $Y = 0$ a $Y = 1$?
9. V předposlední větě v odstavci za vzorcem (4.2) se říká, že derivace integrálu ve vzorci (4.2) je nulová?
10. V práci nejsou průběžně uváděné žádné odkazy na použitou literaturu.

Domnívám se, že uchazeč se zřejmě až příliš soustředil na algebraické úpravy a naprogramování navržených algoritmů a nezbylo mu pak dost času na přehledné uspořádání bakalářské práce. Výsledná práce je bohužel nepříliš čitelná a působí zmateně.

Práci doporučuji uznat jako bakalářskou, pouze pokud uchazeč při obhajobě uspokojivě odpoví na výše uvedené připomínky.

RNDr. Zdeněk Hlávka, Ph.D.