

POSUDEK VEDOUCÍHO DIPLOMOVÉ PRÁCE

Název: Cross-validation and its use in statistics

Autor: Bc. Aibat Kossumov

SHRNUTÍ OBSAHU PRÁCE

V první části práce autor shrnuje využití krosvalidace pro volbu vyhlazovacího parametru pro jádrové odhady hustoty a jádrové odhady regrese. Hlavní část práce pak spočívá ve využití krosvalidace při hledání vhodného lineárního modelu. V předložené práci se přitom autor zaměřil na metody popsané v článku: *J. Shao, Linear model selection by cross-validation. Journal of the American Statistical Association, 88:486–494, 1993.*

Tématem práce bylo představit využití krosvalidačních technik ve statistice a bylo dle mého názoru naplněno.

Vlastní příspěvek autora spočívá vidím v několika oblastech. Část o jádrových odhadech je spíše přehledového charakteru, kde vlastním přínosem autora je pak mnohem pečlivější důkaz Lemmatu 3 než ve zdrojovém článku. V tomto duchu autor pokračoval zejména v hlavní části práce, která se týká lineárních modelů. Zde autor mnohem podrobněji rozepisuje důkazy některých tvrzení z výše zmíněného článku *Shao (1993)*. Řada tvrzení použitá v argumentaci článku totiž není vůbec očividná a některá vyžadují netriviální odvození.

V praktické rovině pak autor využil představené krosvalidační metody v aplikaci, kdy se snažíme na základě průběžných volebních výsledků odhadnout celkové volební výsledky. Tato aplikace je mimo jiné zajímavou ilustrací těžkostí, se kterými se člověk potýká, když se snaží použít teoretické výsledky v reálných aplikacích.

Matematická úroveň. Z matematického hlediska považuji práci ze spíše obtížnější. I když důkazy typicky nejsou původní, tak je autor podrobně rozpracoval. Doplnil řadu dílčích kroků včetně drobnějších tvrzení, která sám dokázal.

Práce se zdroji. Na kolik jsem schopen posoudit, tak všechny použité zdroje jsou řádně citovány.

Formální úprava. Formální úroveň práce je velmi dobrá.

CELKOVÉ HODNOCENÍ PRÁCE

Práci považuji za nadprůměrnou. Autor pracoval velmi svědomitě a samostatně. Na některých místech by se práce dala napsat s větším nadhledem a průhledněji, něco více rozepsat a něco naopak zkrátit a zpřehlednit. Ale to by vzhledem ke složitosti tématu bylo již (dle mého názoru) výrazně nad hranicí toho, co se lze očekávat i od kvalitní diplomové práce.

ZÁVĚR

Domnívám, že práce bez jakékoliv pochybnosti splňuje nároky kladené na diplomovou práci na programu Pravděpodobnost, matematická statistika a ekonometrie a doporučuji ji za ni uznat.

doc. Ing. Marek Omelka, Ph.D.

KPMS MFF UK

31. května 2024