

## POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

**Název:** Cross validation and its use in statistics

**Autor:** Aibat Kossumov

### SHRNUTÍ OBSAHU PRÁCE

Práce pojednává o možném využití tzv. "cross validation" ve statistice. To jsou v principu metody, které vhodně rozdělí data, z nichž jedna část je využita ke konstrukci modelu a druhá část k jeho validaci.

První dvě kapitoly se zaměřují na "cross-validation" při volbě šířky okénka v jádrovém odhadu hustot a jádrové regresi. Teoretické výsledky jsou také odvozeny a provedena simulační studie.

Třetí kapitola (ta je jádrem celé práce) je zaměřena na "cross validation" při volbě lineárního modelu. Studovány jsou 4 možnosti "obecná metoda", "leave-one-out cross-validation", "balanced incomplete cross-validation" a "Monte Carlo cross-validation".

Výsledky třetí kapitoly jsou aplikovány na data z parlamentních a prezidentských voleb v České republice v letech 2021 a 2023. Aplikace je netriviální.

### CELKOVÉ HODNOCENÍ PRÁCE

**Téma práce.** Téma práce považuji za spíše náročnější, ale přiměřené. Zadání práce bylo určité splněno.

**Vlastní příspěvek.** První dvě kapitoly jsou zpracovány na základě citované literatury doplněné vlastními simulacemi a diskuzí. Třetí kapitola sice vychází z článku J. Shao (Linear model selection by cross-validation. Journal of the American Statistical Association, 88:486–494, 1993), ale diplomant ho výrazně přepracoval, doplnil nejen obecný text, ale hlavně pečlivě doplnil většinu důkazů, což považuji v daném problému za dost náročné. Doplnil také různá upřesnění, diskuze a hlavně aplikoval teoretické výsledky na reálná data týkající se voleb v České republice. Připojil velice podrobnou diskuzi.

**Práce se zdroji.** Zdroje jsou správně citovány. V práci nenalezla doslova zkopírované nebo otrocky přeložené pasáže.

**Formální úprava.** Z formálního hlediska je práce vypracovaná velice pečlivě včetně formální stránky. Práce je psána ve velice slušné angličtině.

### PŘIPOMÍNKY A OTÁZKY

- Dotazy: (i) ještě nějakou diskuzi přidat o volbě  $n_c$  a  $n_v$ ,  
(ii) u aplikace na volby uvést podrobně souvislost se značením v teoretické části 3. kapitoly,  
(iii) Doplnit přesnou informaci co je  $M_*$ .
- 
- Drobné poznámky a připomínky:  
(1) str. 32 a další– předpoklady na  $x_i$  by měli být přesnější, mohou být i náhodné. Jak je to v aplikaci v sekci 3.4?

- (2) častější odkazy na zavedené značení, např.  $M_*$ ,
- (3) str.32, pod (3.4) – je nutné, aby počet  $\varepsilon_j$  a  $\tilde{\varepsilon}_i$  byl stejný?

ZÁVĚR Práci považuji za vynikající a doporučuji ji uznat jako diplomou práci.

*Návrh klasifikace vedoucí/oponent sdělí předsedovi zkušební (sub)komise.*

Marie Hušková  
KPMS MFF UK  
1. června 2024