

## POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

**Název:** Výpočet a aplikace MCD estimátoru pro robustní statistické analýzy

**Autor:** Kristýna Sommerová

### **Shrnutí obsahu práce**

Práce se věnuje tématu z oblasti robustní statistiky a to konkrétně MCD estimátoru pro určení střední hodnoty a varianční matice mnohorozměrných dat. Text se zaměřuje především na algoritmus fastMCD široce využívaný k výpočtu aproximace MCD estimátoru. Kapitola 1 shrnuje základní statistické pojmy z dané oblasti, kapitola 2 popisuje samotný (komplikovaný) algoritmus. Kapitola 3 pak studuje některé paměťové a výpočetní náklady algoritmu a komentuje určité otázky jeho stability. Tyto výsledky jsou doplněny Kapitolou 4 ilustrující chování různých variant algoritmu na konkrétních testovacích datech.

### **Celkové hodnocení práce**

**Téma práce.** Samotné téma lze považovat za přiměřeně náročné. Domnívám se však, že v práci bylo pojmuto příliš zešíroka, což jeho zpracování ztížilo. Např. všeobecný úvod do robustní statistiky vedl k nutnosti zavedení řady pojmů, které k jádru práce nebyly potřebné, a i přesto se v Kapitolách 1 a 2 objevují další pojmy, které jsou vágní či nejsou definovány (např. populační střední hodnota). Kapitola 1 pak působí poněkud neorganizovaně. Stručnější přístup by jistě umožnil pečlivější vypracování.

**Vlastní příspěvek.** Za vlastní příspěvek považuji některé výsledky Kapitoly 3, konkrétně vyčíslení a porovnání paměťových nákladů, hlavních výpočetních nákladů a diskusi ohledně stability výpočtu pro fastMCD s různými navrženými variantami výpočtu.

**Matematická úroveň.** Matematická úroveň práce je dobrá. Objevují se některé pojmy nedefinované či definované ne zcela přesně, případně i nepřesná vyjádření, není jich ale vzhledem k délce práce mnoho.

**Práce se zdroji.** Práce obsahuje řadu referencí, postrádám pouze odkazy na literaturu, z níž byly čerpány základní pojmy a fakta v Sekci 1.1, 1.2 a úvod 3.3.

**Formální úprava.** Úprava práce je vhodná. Práce obsahuje menší množství překlepů apod. Bylo by vhodné vyhnout se pojmům jako „BÚNO“ a také používat více konzistentní značení („*det*“ pro determinant se používá někdy se závorkou a někdy bez, náhodný vektor se značí nejprve *ksí* a pak *x*, atd.).

### **Připomínky a otázky**

Připomínky jsou uvedeny výše. Otázky:

1. Jaké hodnoty  $n$  (počet pozorování) jsou obvyklé pro data, pro která se běžně používá MCD estimátor? Překvapilo mne tvrzení, že  $n/300$  je malé (strana 25).
2. V Sekci 3.3.5 se diskutuje, mimo jiné, výpočet  $\det(S_k)$  je-li tato hodnota malá. Jak se po logaritmizaci rozhodne, kterou hodnotu již budeme numericky považovat na nulovou?

### **Závěr**

Práci považuji za velmi dobrou a doporučuji ji uznat jako bakalářskou práci.

Jméno oponenta, podpis RNDr. Iveta Hnětynková, PhD.

Pracoviště KNM MFF UK

Datum 25.8.2016