

POSUDEK DIPLOMOVÉ PRÁCE (POSUDEK VEDOUCÍHO).

Název diplomové práce: Robustní klasifikace a diskriminace

Autor: Dita Rensová

Cílem práce bylo popsat existující robustní metody pro klasifikační a diskriminační analýzu i analýzu hlavních komponent a porovnat jejich chování pomocí numerické studie. Kapitola 1 opakuje klasifikační pravidla, která tvoří klasickou součást mnohorozměrné statistiky. Kapitola 2 představuje podrobně vypracovaný přehled robustních odhadů pro střední hodnotu a varianční matici, které jsou nezbytným nástrojem pro konstrukci robustních metod v následujících kapitolách. Konkrétně kapitola 3 vysvětluje, jak je lze použít při robustní klasifikační analýze a kapitola 4 při různých robustních verzích analýzy hlavních komponent. Chování jednotlivých robustních metod z kapitol 3 a 4 je studováno v kapitole 5 pomocí simulací a v kapitole 6 při analýze dvou reálných datových souborů.

Autorka se seznámila s řadou poznatků, které výrazně přesahují rozsah základních kurzů robustní statistiky i mnohorozměrných statistických metod. Ve skutečnosti již samotné téma robustních odhadů pro mnohorozměrná data přesahuje rámec výuky na MFF UK. Práce je psaná čtivě a srozumitelně a má jasnou strukturu, autorka má vhled do tématu, pojmy a značení zavádí s pečlivostí. Autorka s porozuměním popisuje jednotlivé metody, věnuje se popisu algoritmů pro výpočet odhadů. Je škoda, že se jen stručně věnuje vlastnostem robustních klasifikačních metod v kapitole 3. Za přehledné považuji i programy v softwaru R, které autorka použila při výpočtech a které jsou obsaženy na přiloženém CD.

Za zajímavé považuji výsledky porovnání různých metod v příkladech v kapitole 6.1 a 6.2. V obou případech jde o data obsahující odlehlé hodnoty, a přitom některé robustní odhady vedou k výrazně horším výsledkům než klasické postupy. Zde není snadné najít vysvětlení, protože pro jednotlivé robustní metody není známo, v jaké konkrétní situaci mají být nejvhodnější. Navíc oproti existujícím srovnáním robustních klasifikačních metod (např. v simulacích v článku Todorov a Pires (2007)) diplomová práce i srovnává vliv (klasické i robustní) analýzy hlavních komponent.

Kladně hodnotím, že autorka pracovala velmi samostatně, jak při rešerši v literatuře, tak i při všech výpočtech. Práce má také pěknou formální úpravu.

K práci mám některé drobné připomínky:

- Str. 2. Nesouhlasím s tvrzením, že u mnohorozměrných statistických metod se typicky předpokládá, že známe rozdělení dat a *známe parametry, které toto rozdělení určují*.
- Str. 11. Nevidím zde rozdíl mezi definicí ortogonální ekvivariance a afinní ekvivariance.
- Str. 12. Věta 1 převzatá z článku Davies (1987) by měla předpokládat eliptické rozdělení.

- Str. 16, vzorec (2.15). Vhodnější značení než $d(x; y, A)$ by bylo $d(x, y; A)$.
- Str.30-31. Chyby ve značení, zřejmě má být $\alpha_1, \dots, \alpha_p$ a q se volí tak, aby

$$\frac{\sum_{i=1}^q \lambda_i}{\sigma^2} \geq \alpha.$$

- Str. 48. Z pěti uvažovaných proměnných obsahuje jejich český seznam jen čtyři.
- V některých vzorcích vadí mezera za desetinnou čárkou (např. str. 36, str. 41).
- Práce obsahuje některé překlepy. Jejich počet není vysoký.

Předložená práce velmi úspěšně splnila zadání, a proto doporučuji uznat předloženou práci jako diplomovou práci.

RNDr. Jan Kalina, Ph.D.
Ústav informatiky AV ČR, v.v.i.
V Praze, dne 19. srpna 2013