

POSUDEK VEDOUCÍHO DIPLOMOVÉ PRÁCE

Bc. Lukáš Bandas: Klasifikace na základě longitudinálních pozorování

Práce se zabývá problémem klasifikace na základě longitudinálních pozorování, přičemž se zaměřuje zejména na klasifikaci založenou na pravidle, které je vytvořeno s pomocí „tréninkových dat“, tj. dat, u nichž je známa příslušnost do skupin, do nichž klasifikujeme. Jedná se tedy o zobecnění klasické diskriminační analýzy na případ longitudinálních dat, která nelze typicky reprezentovat pomocí sady nezávislých stejně rozdělených náhodných veličin, resp. vektorů.

V souhrnu se jedná (v souladu se zadáním) o práci kompilační, jejímž cílem bylo prokázat zvládnutí některých oblastí statistiky rozšiřujících standardní studijní plány. Většina v literatuře dostupných postupů pro diskriminační analýzu pro longitudinální data je založena na rozšíření lineárního smíšeného modelu, a proto je tento ve druhé kapitole definován. Ve stejné kapitole jsou dále diskutovány základní vlastnosti tohoto modelu (potřebné v dalších částech práce) a problematika odhadu jeho parametrů. Třetí kapitola začíná shrnutím klasické diskriminační analýzy a pokračuje jejími rozšířeními vhodnými v kontextu longitudinálních dat. Čtvrtá kapitola je věnována aplikacím dříve diskutovaným postupům, které jsou empiricky porovnány pomocí simulační studie v kapitole páté.

Práci by obecně slušelo mnohem více matematické přesnosti. Mnohé formulace (např. „*Subjekt (pacient) \mathbf{Y}_i^g lze nyní v rámci normálního lineárního smíšeného modelu reprezentovat několika náhodnými vektory a jimi souvisejícími rozděleními.*“) nejsou nejšťastnější. Na mnoha místech je autor až příliš stručný, respektive si ne zcela uvědomuje, která fakta a v jaké formě jsou nezávislému čtenáři známa, a která ještě nikoliv. Například úvod k jednotlivým metodám diskriminace začínající v přibližně polovině strany 20 je dle mého názoru pro problematiku neznalého čtenáře velice obtížně pochopitelný. Též z popisu scénáře simulační studie na str. 46 není zcela jasné, jaké scénáře byly vlastně uvažovány.

Taktéž stylisticky by bylo možné práci v mnohém vylepšit. Zejména členění textu je mnohdy nelogické a v rámci jednoho oddílu textu nalezneme dvě jenom velice okrajově související témata. Například část textu na str. 12 začínající „*Z předešlého textu není zcela jasné. . .*“ má pouze velice malou souvislost s tématem oddílu 2.2.2, do kterého je včleněna.

S ohledem na celkový průběh přípravy diplomové práce mohu konstatovat, že diplomant je schopen samostatně nastudovat, pochopit, v potřebné míře rozšířit a následně též aplikovat potřebné teoretické poznatky. Následně má však jisté problémy vše shrnout ve formě matematicky dobře strukturovaného textu. Největší nedostatek shledávám v tom, že diplomantovi činí potíže dobře si utřídit, která fakta již byla uvedena v dřívějších částech textu (a tudíž není nutné je opakovat),

která fakta čtenář ještě nezná (a tudíž je potřeba je uvést), resp. správně odlišit méně podstatné od podstatného.

Přes všechny uvedené nedostatky práce dle mého názoru **splňuje** požadavky kladené na diplomovou práci pro studijní obor Pravděpodobnost, matematická statistika a ekonometrie studijního programu Matematika na Matematicko-fyzikální fakultě Univerzity Karlovy v Praze a **doporučuji** ji k obhajobě.

V Praze dne 3. května 2012

RNDr. Arnošt Komárek, Ph.D.
vedoucí diplomové práce