

## POSUDEK VEDOUcíHO BAKALÁŘSKÉ PRÁCE

**Název:** Klasifikace založená na směsových modelech

**Autor:** Lucie Janečková

### SHRNUTÍ OBSAHU PRÁCE

Práce je věnována problematice směsí konečných rozdělání a jejich možnému využití ke klasifikaci. V první kapitole je zaveden pojem konečné směsi rozdělání, který je dále zúžen na směs normálních rozdělání. Dále je zde představena souvislost konečné směsi rozdělání s klasifikačním problémem. Druhá kapitola je věnována odhadu parametrů směsi metodou maximální věrohodnosti (ML). Kromě vlastního odhadu je zde diskutována též omezenost věrohodnostní funkce a s tím související existence ML odhadu. Ve třetí kapitole se autorka věnuje tzv. EM algoritmu, který je obvykle využíván k praktickému nalezení ML odhadu parametrů konečné směsi. Čtvrtá kapitola aplikuje studované postupy při analýze reálných archeologických dat.

Je vhodné též poznamenat, že se jedná o přepracování dříve neobhájené práce, kdy významných změn doznaly zejména třetí a čtvrtá kapitola práce.

### CELKOVÉ HODNOCENÍ PRÁCE

**Téma práce.** Cílem práce bylo seznámit se s problematikou konečných směsí rozdělání a jejich využitím při klasifikaci. Z povahy zadání se jednalo o téma spíše kompilační. Cílem vypracování bakalářské práce bylo rozšířit si znalosti získané v základních kurzech statistiky, tyto shrnout ve formě matematického textu a ukázat schopnost vše prakticky aplikovat na reálných datech. Zadání práce bylo dle názoru vedoucího práce splněno.

**Vlastní příspěvek.** Autorka přehledně shrnula studované téma. Vlastnosti věrohodnostní funkce zformulované ve formě tvrzení 2 a věty 3 jsou sice dobře známé, ale příslušné důkazy/odvození se v literatuře nalézají velmi nesnadno. Pro účely bakalářské práce provedla proto autorka svoje vlastní odvození, která uvádí ve druhé kapitole. Autorka též samostatně zanalyzovala reálná data (viz kapitola 4). Oproti předchozí verzi práce, ve které byla tato kapitola ve značně neuspokojivém stavu došlo ke zlepšení.

**Matematická úroveň.** Matematické části textu jsou uváděny ve formě definic a tvrzení při konzistentním používání zavedeného značení. Hrubé chyby, mohu-li soudit, se v práci nevyskytují. Autorka se nicméně nevyhla některým nepřesnostem. Např. v oddíle 4.2 se „skutečnými hodnotami neznámých parametrů“ zřejmě myslí odhady těchto parametrů při známé příslušnosti do jedné ze skupin. Obdobně vektory  $\mu_1, \dots, \mu_3$  ve vzorci uprostřed strany 18 mají označovat tyto odhady, tj. příslušné skupinové výběrové průměry.

**Práce se zdroji.** Zdroje, z nichž práce vychází, jsou řádně a též formálně správným způsobem citovány.

**Formální úprava.** Formální úprava práce je na přiměřené úrovni. Zjevné překlepy nebo chyby se v práci vyskytují v míře malé a odpovídající povaze a rozsahu práce.

## ZÁVĚR

Celkově se jedná o průměrnou práci. Zejména čtvrtá kapitola, jež i v letošním roce vznikala z velké části až během července a s jenom minimálním vstupem ze strany vedoucího práce, by i nadále snesla jistá zlepšení. Oproti dříve neobhájené variantě nicméně došlo k významnému kvalitativnímu posunu. Práce jako celek dle mého názoru splňuje požadavky kladené na bakalářskou práci na oboru *Obecná matematika* a za takovou ji též **doporučuji** uznat.

doc. RNDr. Arnošt Komárek, Ph.D.

Katedra pravděpodobnosti a matematické statistiky  
Matematicko-fyzikální fakulta Univerzity Karlovy

V Jindřichově Hradci dne 24. července 2023