

Posudek diplomové práce **Martiny Janáčkové,**

studijní obor Aplikace matematiky v ekonomii,
KMAaAM PŘF UP Olomouc, školní rok 2012/2013

Statistické modely a neuronové sítě

Diplomová práce Martiny Janáčkové na téma Statistické modely a neuronové sítě má celkem 87 stran, je rozčleněna do čtyř kapitol a obsahuje rovněž dvě přílohy. Autorka čerpala z celkem 40 zdrojů, na které odkazuje v seznamu literatury.

První až třetí kapitola zahrnuje teoretickou část práce. Autorka postupně představuje tři metody, které lze využít pro úlohy klasifikačního typu. Nejprve se podrobně věnuje neuronovým sítím, a to včetně jejich biologické inspirace. Poté v kapitole druhé představuje rozhodovací stromy (včetně tzv. soft tree) a celou kolekci završuje krátkou připomínkou logistické regrese. Každý popisovaný model je aplikován na data, která se týkají rozpoznávání malignity buněk.

Podstatnou část práce tvoří čtvrtá kapitola. V ní jsou jednotlivé modely, označované jako ANN (umělé neuronové sítě), TREE (klasifikační stromy), LR (logistická regrese), vzájemně porovnány. Autorka využívá parametrickou analýzu rozptylu, neparametrický Kruskalův-Wallisův test a Hotellingovu statistiku. Každý z testovacích postupů stručně popisuje, včetně využití softwaru R. Samotné porovnání je realizováno na datech reprezentujících 3×100 výsledků (procento špatně zařazených pozorování z testovací podmnožiny původních dat). Výsledky testování jsou vždy komentovány.

Práce je poměrně rozsáhlá, co do počtu stran i šíře tématu. Neuronové sítě a klasifikační stromy nejsou vůbec součástí standardní výuky oboru a logistická regrese je probírána pouze okrajově. Oceňuji autorčin přístup k vlastnímu tématu, kdy musela samostatně dohledat zdrojovou literaturu a s jednotlivými postupy se seznámit. Naplnila tak zadání práce, včetně využití dostupného softwaru (software R a jeho doplňkové balíčky). Sama vytvořila funkce, které v používaných balíčcích chyběly (párové porovnání v případě signifikantního Kruskalova-Wallisova testu, aplikace umělé neuronové sítě vytvořené funkcí `neuralnet()` na testovací množinu dat). Součástí práce je rovněž univerzální algoritmus provádějící 100 opakování rozdělení dat na trénovací a testovací množinu a aplikaci všech tří metod, jehož výstupem je již zmiňovaná matice výsledků.

Bohužel jsem v práci našla řadu překlepů (7_3 , 9^2 , 22_6 , 22_{12} , 23_7 , 31^8 , 32_1 , 38_5 , 40_1 , 43_8 , 44_2 , 52^4 , 54^6 , 54_4 , 56_1 , 59^5 , 61_7 , 63^1 , 64^9 , 66^4 , 67^{10} , 71_1 , 74_4 , 76^5 , 76^{16} , 77_2) a nepřesných formulací (např. 31^{10-11} , 42_1 , 52^1 , tab. 8 na str. 55, 61^{3-8} , 76_13). Orientaci v textu čtenáři sťažují chyby v číslování tabulek a neseřazený seznam literatury. Nedostatky však nejsou závažného charakteru, spíše mají charakter překlepů vzniklých z nepozornosti v závěrečné fázi úpravy textu. Diplomovou práci hodnotím jako kvalitní, doporučuji ji k obhajobě a v případě jejího kvalitního průběhu, kdy budou některé formulace uvedeny na pravou míru, navrhuji hodnocení **stupněm B**.

V Olomouci dne 23. 1. 2013

Mgr. Jana Vrbková, Ph.D.
vedoucí práce