

Filip Šimsa

Kritéria těsnosti regrese dle typu vysvětlované proměnné
posudek oponenta bakalářské práce

Student se v práci zabývá měřítky kvality regresního modelu, konkrétně modelu logistické regrese a jejího zobecnění na multinomickou odezvu. Toto téma je podle mého názoru na bakalářskou práci docela obtížné a byl jsem zvědav, jak si s ním student poradí.

Musím říci, že práce je na bakalářskou práci docela obsáhlá. Bohužel mi z ní není úplně jasné, co je přesně vlastním přínosem autora. Toto by se mělo v průběhu obhajoby vyjasnit.

Pokud jde o jazykovou úroveň textu, bylo by co vylepšovat. V češtině není zvykem psát názvy jako *vysvětlovaný součet čtverců* (strana 5) s velkým počátečním písmenem, podobně nazývat Coxův-Snellův koeficient jako *Cox and Snellův* (strana 13) nevypadá dobře, stejně jako ve vzorcích označovat Giniho koeficient jako *Gini*.

Matematicky by si práce zasloužila také upravit. Vedle zvláštních interpretací, například v regresním modelu je $\hat{Y} = X\beta$ označeno za odhad náhodné veličiny Y (strana 5), některých překlepů (namátkově na řádku 15, strana 29, řádek 10 na straně 8), mi nejvíce vadí, že práce je takovým vyprávěním s nedostatečnou úrovní matematického formalismu. Označit S_n symetrickou grupu permutací na straně 18 dole a hned v následujících vzorcích na straně 19 dát S_m úplně jiný význam je nešťastné. V práci není žádná formální definice ani tvrzení s důkazem. Přitom je mnoho míst, kde by se nějaký důkaz přímo nabízel. Neustále se hovoří o přibližných vztazích (značeno \approx), aniž by bylo jasné, proč tyto přibližné vztahy platí, v jakém smyslu (a jak moc přibližně) a jaký z toho máme vyvodit závěr. Celkově tedy nevím, co si z této práce mám odnést za hlavní poučení a jak mne matematicky obohatí. Proč například na straně 6 hovoříme o tom, že Y_i má alternativní rozdělení s parametrem p_i , když o pár řádků níže odhadujeme $E(Y_i|\mathbf{x}_i) = \pi(\mathbf{x}_i)$ a jak zdůvodníme použití logistické regrese v (1.10)? Jaký je vlastně vztah mezi p_i a $\pi(\mathbf{x}_i)$? A proč je v práci zavedena Kolmogorovova-Smirnovova statistika? Dá se nějak využít například k testování submodelu?

Podle mého by bylo lepší, kdyby se student v práci zaměřil na méně koeficientů a spíš věnoval pozornost formálnímu zpracování textu. Přesto si myslím, že svou práci splnil zadání bakalářské práce, a proto doporučuji **uznat jeho práci jako bakalářskou**.

Daniel Hlubinka
28.8.2012