Posudek vedoucího diplomové práce

Název: Post-selection Inference: Lasso & Group Lasso
Autor: Bc. Vojtěch Bouř

Shrnutí obsahu práce

Autor sa vo svojej diplomovej práci venuje pomerne pokročilému spôsobu odhadovania koeficientov v obyčajnom lineárnom regresnom modeli, pričom využíva moderné postupy z oblasti atomického hľadania cieľového odhadu pomocou tzv. LASSO regularizácie a uplatňuje najnovšie výsledky z teórie pravdepodobnosti a štatistiky ohľadom inferencie v takto skonštruovaných modeloch. Tématické zameranie práce je pomerne náročné a vyžaduje podrobné naštudovanie a pochopenie nie malého, a pritom pomerne zložitého aparátu rôznych štatistických metód a postupov, pomocou ktorých je možné robiť štatisticky korektné závery a všeobecné platné tvrdenia v LASSO regularizovaných regresných modeloch.

Tieto modely vznikajú kombináciou klasickej L_2 stratovej funkcie (bežnej v obyčajnej lineárnej regresii) a tzv. LASSO penalty (t.j., L_1 normy z príslušných odhadov). Ich základnou výhodou je, že umožňujú automatický (ale jednoduchý a intuitívny) výber relevantných premenných. Túto vlastnosť si pritom zachovávajú aj v prípadoch, keď je počet sledovaných premenných výrazne vyšší, ako je počet subjektov, resp. pozorovaní (zatiaľ čo klasické regresné modely v takýchto situáciach zlyhávajú). Na druhej strane, základná nevýhoda spočíva práve v náhodnom výbere spomínaných premenných (tzv. random selection). Túto náhodnosť vo výbere premenných klasické inferenčné metódy používané v štandardných regresných modeloch neberú do úvahy a preto nie su korektné (cca pred troma rokmi dokonca neboli známe žiadne teoretické postupy, ktoré by umožňovali platnú inferenciu v LASSO regularizovaných modeloch).

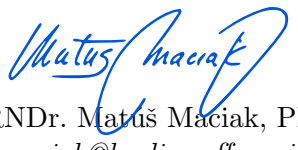
Autor vo svojej práci popisuje najnovšie výsledky, ktoré sa snažia s náhodným výberom premenných v LASSO modeloch vysporiadať a poskytnúť štatistikom platné inferenčné postupy. V práci sú prehľadne uvedené všetky momentálne známe štatistické testy, z ktorých niektoré fungujú na výrazne rozdielných princípoch, a autor podrobne popisuje spôsob ich aplikácie v reálnych situáciach.

Zaujímavým a celkom cenným príspevkom autora je už samotné teoretické porovnanie týchto testov, nakoľko sa jedná o nové a málo preskúmané postupy. Zaujimavý je aj autorov pokus o návrhnutie štatistického testu, ktorý by bolo možne aplikovať na skupinové LASSO (tzv. group LASSO).

Za hlavný autorov prínos ale považujem zaujímavú, celkom prehľadnú a samostatne vypracovanú simulačnú štúdiu. V tejto simulačnej štúdii autor porovnáva uvedené štatistické testy určené pre LASSO regularizované modely a na niekoľkých názorných príkladoch prezentuje ich empirické vlastnosti z viacerých uhlov pohľadu. Takéto zrovnanie momentálne známych testov určených pre LASSO inferenciu sa dá považovať, podľa môjho názoru, za celkom jedinečné.

Celkovo hodnotim prácu ako pomerne náročnú, ale kvalitne vypracovanú. Autor prejavil samostatné schopnosti, dokázal pomerne zložité teoretické metódy z rôznych zdrojov naštudovať, pochopiť a prezentovať ich ako jednotný celok. Tieto metódy a ich teoretické vlastnosti navyše doplnil o celkom zaujímavú a jedinečnú simulačnú štúdiu.

Prácu doporučujem uznať ako diplomovú prácu.



V Prahe, 27.04.2017

RNDr. Matuš Maciak, Ph.D.
maciak@karlin.mff.cuni.cz