

Oponentský posudek na diplomovou práci:

JOEL JANČAŘÍK: BAYESOVSKÝ VÝBĚR PROMĚNNÝCH

Uchazeč se v předložené práci zabývá úlohou výběru proměnných v lineárním modelu z pohledu Bayesovské statistiky. Představené metody jsou rozděleny na metody založené na indikátoru (kapitola 4) a na metody založené na srážení (shrinkage, kapitola 5). Přínosem práce je zejména popis odvození potřebných podmíněných aposteriorních rozdělání (pro metody navržené v aktuální odborné literatuře) a popis Gibbsova algoritmu.

Struktura práce je logická a téma je zpracováno vyčerpávajícím způsobem. Kvalitu celé práce bohužel snižuje značné množství překlepů, např. na straně 1: překlep v definici průměru, chybějící n ve slově desetinou, chybně rozdělené slovo podm-nožin. Další překlepy viz např. druhá věta v kapitole 1 nebo první věta v kapitole 2.

Další připomínky:

str. 10 Prosím o vysvětlení symbolu $K(z_1, z_2)$ ve vzorci na posledním řádku.

str. 17 Chybí – v nečíslovaných vzorcích před (2.4) a (2.5). Nechybí σ^2 ve vzorci pro Σ^{-1} (za (2.5), viz také druhý řádek na str. 18)?

str. 24 Nevypadlo σ^2 ze vzorců pro podmíněné rozdělání $\beta_d|\delta$ (viz též strana 26)?

str. 27 Překlep v prvním nečíslovaném vzorci v odstavci pro “Podmíněné rozdělání δ_j ”.

str. 28 Prosím o vysvětlení uvedených parametrů inverzního gamma rozdělání (čtvrtý řádek odspodu).

4.2.2 Popisovaná metoda zřejmě silně závisí na volbě měřítka vysvětlujících proměnných. Uvedené interpretaci ale neodpovídá Definice 15.

str. 33 Co přesně se myslí podmínkou $\beta_j \in (-x, x)$ (čtvrtý řádek ve čtvrtém odstavci)? Je β_j konstanta nebo náhodná veličina? Následující odvození opět silně závisí na měřítku vysvětlujících proměnných.

str. 36 Kritérium $0 \in (\beta_j^L, \beta_j^U)$ nemusí fungovat v přítomnosti multikolinearity.

str. 46, poslední řádek Nesouhlasím s popisem výsledků v příloze 6.3. Proč jsou boxploty pro různé hodnoty parametru β_j fakticky stejné? Nemělo by aposteriorní rozdělání odhadovaného parametru záviset na odhadovaném parametru?

str. 50, druhý odstavec Nebylo by možné pracovat s transformovanými vysvětlujícími proměnnými místo jejich vynechání?

V předložené práci postrádám porovnání s nebayesovským přístupem (stačilo by alespoň porovnání výsledků při analýze skutečných dat). Vůbec se nekomentují praktické problémy, které by mohly být způsobené multikolinearitou nebo změnou měřítka vysvětlujících proměnných. Domnívám se, že představené metody nebudou v přítomnosti multikolinearity použitelné.

Přínosem práce je zejména přehledné představení Bayesovského přístupu k výběru proměnných podle aktuální odborné literatury. I přes výše uvedené výhrady doporučuji předloženou práci uznat jako diplomovou.

Doc. RNDr. Zdeněk Hlávka, Ph.D.