

POSUDEK VEDOUCÍHO DIPLOMOVÉ PRÁCE

Název: Bayesovský výběr proměnných

Autor: Joel Jančařík

SHRNUTÍ OBSAHU PRÁCE

Výběrem proměnných v rámci statistické analýzy dat se typicky rozumí výběr regresorů z potenciálně velké množiny proměnných, které mají být zahrnuty v regresním modelu. Předložená práce se zabývá bayesovskými přístupy k této úloze. Práce je členěna do šesti kapitol. Po krátké úvodní kapitole následuje přípravná kapitola, která zavádí potřebné pojmy a shrnuje znalosti z bayesovské statistiky a taktéž teorie *Markov chain Monte Carlo (MCMC)* metod používané v hlavní části práce. Třetí kapitola matematicky zavádí úlohu výběru proměnných s důrazem na výběr proměnných v kontextu lineárního modelu. Hlavní část práce se nachází v kapitolách 4 a 5, které jsou postupně věnovány dvěma třídám metod: výběru proměnných za pomoci indikátoru a metodám založeným na tzv. srážení (*shrinkage*) regresních parametrů. Diskutované metody jsou použity při analýze reálných dat i v rámci simulační studie v kapitole 6.

CELKOVÉ HODNOCENÍ PRÁCE

Téma práce. Bayesovský výběr proměnných je oblastí aktivního výzkumu, jehož hlavní rozvoj započal před přibližně 15 lety. Většina poznatků je k dispozici pouze v roztržštěné formě ve formě časopiseckých článků. Hlavním cílem práce bylo shrnout vybrané metody a doodvodit vybrané vlastnosti, které jsou v člancích uváděny typicky pouze polovičatě, případně vůbec. Tento cíl byl, dle mého názoru, naplněn.

Vlastní příspěvek. Diplomant při jednotném značení shrnul poměrně obsáhlou oblast, pro niž fakticky existují pouze časopisecké prameny. Pro všechny uváděné metody kromě jiného odvodil tzv. plně podmíněná rozdělení, jež jsou potřebná pro implementaci Gibbsova algoritmu a praktické použití uvažovaných metod. Tyto implementoval v jazyce R a následně použil jak na analýze reálných dat, tak v rámci simulační studie.

Matematická úroveň. Práce je psána standardním matematickým stylem. Značení je v celé práci jednotné, tvrzení vycházejí z předem uvedených předpokladů. Práce obsahuje celou řadu vlastních odvození, zejména tzv. plně podmíněných rozdělení. Dále je např. dokázána nevlastnost aposteriorního rozdělení v případě, že se v kontextu normálního lineárního modelu použije přímočaře se nabízející Jeffreysova apriorní hustota pro hyperparametry udávající apriorní přesnosti regresních koeficientů (viz oddíl 5.2).

Faktická správnost. V práci se dle mého názoru nevyskytují závažné faktické chyby. Nicméně, i v konečné verzi práce zůstala i v několika vzorcích drobná opomenutí, respektive překlepy, které hodnotu práce snižují.

Formální úprava a práce se zdroji. Po formální stránce je práce na slušné úrovni až na několik zjevných překlepů, kterým se jistě bylo možné vyhnout. Použité zdroje jsou řádně a obvyklým způsobem citovány.

ZÁVĚR

Práci **doporučuji** uznat jako diplomovou práci.

doc. RNDr. Arnošt Komárek, Ph.D.

Katedra pravděpodobnosti a matematické statistiky
Matematicko-fyzikální fakulta Univerzity Karlovy

V Praze 5.6.2017