

Modelově založené shlukování vícerozměrných longitudinálních dat smíšeného typu

Jan Vávra

3. října 2022

Abstrakt

Mnoho dnešních studií sbírá data opakovaně na těch samých jedincích po předem vymezenou časovou dobu. Takto vzniklá longitudinální data jsou navíc často tvořena číselnými, číselnými, binárními, ordinálními nebo obecně kategoriálními hodnotami. Je zde navrženo několik variant statistických modelů schopných modelovat takováto často velmi korelovaná data sdruženě. Metodologie modelově založeného shlukování je zde použita pro odhalení skryté heterogenity v datech tím, že jedince roztrídí do několika skupin specifických vlastností. Generativní model je zde vytvořen za bayesovského přístupu a jsou zde vyvinuty MCMC metody pro jeho odhad. Vlastnosti stvořených odhadů jsou podrobeny simulační studii. Vyvinutá metodologie je aplikovaná na problémy z reálného prostředí, např. data z lékařské studie o pacientech trpících primární biliární cholangitidou (PBC) či ekonomický dataset o tisících českých domácnostech sledovaných od roku 2005 (databáze EU-SILC).