

BME VIK Számítástudományi és Információelméleti Tanszék

Diplomamunka, önálló labor és TDK témajavaslat

Témakiíró:	Dr. Csáji Balázs Csanád
Beosztás:	tudományos főmunkatárs
Munkahely:	MTA SZTAKI
Email:	balazs.csaji@sztaki.mta.hu
Cím:	Regressziós modellek bizonytalansága
Leírás:	<p>A regresszió – azaz (statikus vagy dinamikus) modellek illesztése zajos megfigyelési adatokra – egy klasszikus statisztikai probléma, amely alapvető fontossággal bír számos területen, például a gépi tanulásban, rendszer identifikációban, jelfeldolgozásban és pénzügyi matematikában. A standard módszerek általában egy pontbecslést szolgáltatnak, azaz egy konkrét modellt választanak ki a megadott modellosztályból egy adott kritérium és a megfigyelési adatok alapján. Azonban gyakorlati szempontból kritikus kérdés az is, hogy felmérjük: mennyire bízhatunk a kapott modellben? Egy lehetséges válasz erre a kérdésre, ha megadott valószínűségű konfidencia halmazokat építünk a pontbecslésünk (a kiválasztott modell) köré. Ezen konfidencia halmazok kiemelkedően fontosak, például, kockázatelemzéséhez és a becslt modellen alapuló megoldások robusztussá tételéhez.</p> <p>A tipikus konfidencia halmaz konstrukciók azonban vagy egy konkrét zaj eloszlást feltételeznek, vagy aszimptotikus eredményekre (a használt pontbecslés határeloszlására) építenek. Ezek a feltételezések viszont korlátozzák a gyakorlati alkalmazhatóságukat.</p> <p>Egy új kutatási irányt jelentenek a nemparametrikus statisztikai módszereken alapuló, eloszlás-független, nem-aszimptotikus becslési megközelítések, amelyek garantált, bizonyos feltételek mellett egzakt, konfidencia halmazokat szolgáltatnak megadott pontbecslések (pl., legkisebb négyzetek, legkisebb abszolút eltérés vagy instrumentális változók) köré. Ilyen módszerek például az LSCR (Leave-out Sign-dominant Correlation Regions), SPS (Sign-Perturbed Sums) és ScoPe (Score Permutation) algoritmusok.</p> <p>A feldolgozandó téma ezen módszerek valamelyikének megértése és alkalmazása valós problémákra; esetlegesen a módszerek kiterjesztése, például, előrejelzési tartományok készítésére, kísérlettervezésre, vagy a megfigyelt rendszer megváltozásának felismerésére. A numerikus kísérletek elvégzéséhez Matlab ismerete ajánlott.</p> <p>Irodalom:</p> <ul style="list-style-type: none">- Csáji, B. Cs.; Campi, M. C.; Weyer, E.: Sign-Perturbed Sums: A New System Identification Approach for Constructing Exact Non-Asymptotic Confidence Regions in Linear Regression Models, IEEE Transactions on Signal Processing, Vol. 69, 2015, pp. 169–181- Weyer, E.; Campi, M. C.; Csáji, B. Cs.: Asymptotic Properties of SPS Confidence Regions, Automatica, Elsevier, Vol. 82, 2017, pp. 287–294- Csáji, B. Cs.: Score Permutation Based Finite Sample Inference for Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) Models, 19th International Conf. on Artificial Intelligence and Statistics (AISTATS), Cadiz, Spain, 2016, pp. 296-304- Campi, M. C.; Csáji, B. Cs.; Garatti, S.; Weyer, E.: Certified System Identification: Towards Distribution-Free Results, 16th IFAC Symposium on System Identification (SYSID 2012), Brussels, Belgium, July 11–13, 2012, pp. 245–255 (Plenary Talk)