

DOKTORANTŪROS STUDIJŲ DALYKO SANDAS

Dalyko pavadinimas	Mokslų kryptis (šaka) kodas	Fakultetas	Katedra
Netiesiniai statistikos modeliai masinių duomenų analizėje	Informatika (09 P)	MIF	Duomenų mokslų ir skaitmeninių technologijų institutas
Studijų būdas	Kreditų skaičius ECTS	Studijų būdas	Kreditų skaičius
Paskaitos (pavasario sem.) individualus	1 4	konsultacijos Seminarai (pavasario sem.)	1 1

Dalyko anotacija

Dalyko tikslas – papildyti studentų turimas žinias apie automatinį (mašininį) mokymąsi netiesinio statistinio modeliavimo žiniomis akcentuojant kritinį statistinį mąstymą.

Kurso kontekste didieji duomenys (angl. *Big data*) suprantami, kaip masiškai be konkretaus tikslo surinkti duomenys ir yra vadinami masiniais duomenimis. Pastarieji duomenys, kaip taisyklė, yra heterogeniški nuo paprasčiausių mažų tekstinių įrašų iki akcijų informacijos kas minutę ar viso genomo duomenų. Šiuos duomenis galima analizuoti traktuojant juos kaip juodąją dėžę (angl. *Black box*) ir naudoti automatinio (mašininio) mokymosi metodus. O galima taikyti statistinį modeliavimą ir, atsižvelgiant į duomenų generavimo procesą, daryti prielaidas apie duomenų skirstinį ir jas testuoti. Kuo konkretnės (griežtesnės) prielaidos dera su duomenimis, tuo turiningesnė ir subtilesnė gautų rezultatų interpretacija ir išvados. Todėl automatinio (mašininio) mokymosi metodus prasminga derinti su netiesinio (parametriniais ir neparametriniais) statistinio modeliavimo metodais.

Pagrindinės temos:

- Klasikiniai ir robastiniai tiesiniai metodai (Agresti)
- Apibendrintieji tiesiniai (AT) modeliai, modelio pasirinkimas ir statistinės išvados (Agresti)
- Pakartotiniai stebiniai, atsitiktiniai veiksniai ir ilgalaikių stebėjimų duomenys (Faraway)
- Mišriųjų veiksnų modeliai, nenormalusis atsako kintamasis (Faraway)
- Bajeso daugialygis (hierarchinis) modeliavimas, Markovo grandinių Monte Karlo metodai (Madigan)
- AT modelio regularizavimas (Agresti)
- Neparametrinė regresija (Faraway), apibendrintieji adityvūs modeliai ir glodinimo metodai (Agresti, Faraway)

Studentų gebėjimai ir atsiskaitymas: Kurso metu studentai, išanalizavę skirtą literatūrą, pasirinktiems duomenims realizuoja apibendrintuosius tiesinius ir neparametrinius regresijos modelius. Esant poreikiui, masinių duomenų analizei moka pritaikyti regularizacijos metodus, sudėtingų struktūrų tyrime – Bajeso metodus. Kursą išklauses studentas suvokia ir geba įvertinti duomenų ir jų analizės rezultatų neapibrėžtumą ir patikimumą. Metodai realizuojami R kalba arba kita studentui ir dėstytojui abipusei priimtina programine įranga. Kursas baigiamas pritaikytų metodų aprašymu (idealiu atveju mokslinė publikacija).

Pagrindinė literatūra

Agresti, A., 2015. Foundations of linear and generalized linear models. John Wiley & Sons.
<http://bayanbox.ir/view/7443147326514856944/Foundations-of-Linear-and-Generalized-Linear-Models-Wiley-Series-in-Probability-and-Statistics-Alan-Agresti-2015.pdf>

Faraway, J.J., 2016. Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models (Vol. 124). CRC press.

Madigan, D., Ridgeway G., 2002. Bayesian Data Analysis for Data Mining.
<https://pdfs.semanticscholar.org/58be/87292ce8c4a0eefca6dd5430368f4af4e177.pdf>

Papildoma literatūra

Alpaydin, E., 2010. Introduction to Machine Learning. The MIT Press. ISBN-10: 0-262-01243-X, ISBN-13: 978-0-262-01243-0.
http://cs.du.edu/~mitchell/mario_books/Introduction_to_Machine_Learning_-_2e_-_Ethem_Alpaydin.pdf

Han, J., Kamber, M., 2006. Data Mining Concepts and Techniques. 2nd ed. CA: Morgan Kaufmann Publishers is an imprint of Elsevier.
http://ccs1.hnue.edu.vn/hungtd/DM2012/DataMining_BOOK.pdf
<https://pdfs.semanticscholar.org/02e0/bc77460469aefec5bd794ee6c4efc15e6adb.pdf>

Konsultuojančiųjų dėstytojų vardas, pavardė	Mokslo laipsnis	Svarbiausieji darbai mokslo kryptyje (šakoje) paskelbti per pastaruosius 5 metus
Audronė Jakaitienė	dr.	<p>Jakaitienė, A., Avino, M. and Guarracino, M.R. (2017). Beta-Binomial Model for the Detection of Rare Mutations in Pooled Next-Generation Sequencing Experiments. <i>Journal of Computational Biology</i>, Vol. 24 No. 4, p. 357-367. DOI:10.1089/cmb.2016.0106.</p> <p>Pranckienė, L., Jakaitienė, A. and Kucinskas, V. (2016). An Evaluation of De Novo Mutation Rate in Lithuanian Exome. <i>Human Heredity</i>. ISSN: 0001-5652, Vol. 81, No. 4, pp. 234-235.</p> <p>Ginevičienė V., Jakaitienė A., Pranculis A., Milašius K., Tubelis L., Utkus A. (2014). AMPD1 rs17602729 is associated with physical performance of sprint and power in elite Lithuanian athletes. <i>BMC Genetics</i> 15:58.</p> <p>Jakaitienė, A., Klyviene, V. (2013). "Tax Elasticities – Factors Causing Fluctuations in the Short and Long Run, the Case of Lithuania", <i>Transformations in Business & Economics</i>, Vol. 12, No 1 (28), pp.42-59.</p> <p>Jakaitienė, A., Dees, S., (2012). Forecasting the world economy in the short-term. <i>The World Economy</i>, Volume 35, Issue 3, p. 331–350.</p>
Marijus Radavičius	dr.	<p>Giedrė Lapinskienė, Kęstutis Peleckis, Marijus Radavičius (2015) Economic development and greenhouse gas emissions in the European Union countries. <i>Journal of Business Economics and Management</i>, 16 (6): 1131-1145.</p> <p>Piaseckienė, K.; Radavičius, M. (2014). Empirical Bayes estimators of structural distribution of words in Lithuanian texts. <i>Nonlinear Analysis: Modelling and Control</i>, 19 (4), 611-625. ISSN: 1392-5113.</p> <p>Matulevičienė, A., Preikšaitienė, E., Linkevičienė, L., Radavičius, M., Molytė, A., Utkus, A., Kučinskas, V. (2013). Heterogeneity of Oral Clefts in Relation to Associated Congenital Anomalies, <i>Medicina</i>, 49(2): 61-66. ISSN 1010-660X.</p> <p>Ričardas Taraškevičius, Rimantė Zinkutė, Rimutė Stakėnienė, and Marijus Radavičius (2013) Case Study of the Relationship between Aqua Regia and Real Total Contents of Harmful Trace Elements in Some European Soils. <i>Journal of Chemistry</i>, vol. 2013, Article ID 678140, 15 pages. doi:10.1155/2013/678140.</p>
Olga Kurasova	dr.	<p>Medvedev, V.; Kurasova, O.; Bernatavičienė, J.; Treigys, P.; Marcinkevičius, V.; Dzemyda, G. A new web-based solution for modelling data mining processes. <i>Simulation Modelling Practice and Theory</i>. Amsterdam : Elsevier Science. ISSN</p>

		1569-190X. eISSN 1878-1462. 2017, Vol. 76, p. 34-46. DOI: 10.1016/j.simpat.2017.03.001.
		Bernatavičienė, J.; Dzemyda, G.; Kurasova, O.; Marcinkevičius, V.; Medvedev, V.; Treigys, P. Cloud computing approach for intelligent visualization of multidimensional data. Advances in stochastic and deterministic global optimization. Ser.: Springer optimization and its applications. ISSN 1931-6828. Vol. 107. Cham: Springer International Publishing Switzerland, 2016. ISBN 9783319299730. p. 73-85. DOI: 10.1007/978-3-319-29975-4_5.
		Orts, F.; Filatovas, E.; Ortega, G.; Kurasova, O.; Garzon, E. M. Improving the energy efficiency of SMACOF for multidimensional scaling on modern architectures. Journal of Supercomputing. Dordrecht: Springer. ISSN 0920-8542. eISSN 1573-0484. 2018, First published online, [13 p.]. DOI: 10.1007/s11227-018-2285-x.