

DOKTORANTŪROS STUDIJŲ DALYKO SANDAS

Dalyko pavadinimas	Mokslo kryptis (šaka) kodas	Fakultetas	Katedra
Daugiamatė statistika, laiko eilutės	09 P	MIF	Informatikos
Studijų būdas	Kreditų skaičius	Studijų būdas	Kreditų skaičius
paskaitos	4	konsultacijos	1
individualus		seminarai	2

Dalyko anotacija. Supažindinama su pagrindiniais matematinės statistikos metodais, skirtais daugiamačiams duomenims ir laiko eilutėms aprašyti, juos analizuoti, klasifikavimo, prognozavimo taisyklėms sudaryti, jų tikslumui įvertinti.

1. Prielaidos apie matavimų atsitiktinį pobūdį ir jų kritinis vertinimas. Imties sudarymo principai. Atsitiktinių dydžių charakterizavimas vidurkiu, dispersija. Koreliacija tarp dviejų atsitiktinių dydžių. Vienamačių ir daugiamačių atsitiktinių dydžių aprašymas pasiskirstymo funkcija. Vienamačiai ir daugiamačiai pasiskirstymo tankiai. Laiko eilutės. Vienamatis ir daugiamatis normaliniai (Gauso) pasiskirstymo tankiai. Jų skaitinės charakteristikos.
2. Dvimačio pasiskirstymo tankio vaizdavimas plokštumoje. Sąlyginis pasiskirstymo tankis. Sąlyginis vidurkis. Regresijos lygties sąvoka. Jos interpretavimas ir aiškinimas sąlyginio vidurkio pagalba. Daugybinė regresija. Daugybinės regresijos lygties gavimas minimalių kvadratų metodu. Prognozavimo tikslumo sąvoka, jo charakterizavimas vidutine kvadratine paklaida bei daugybinės koreliacijos koeficientu. Vienmačių bei daugiamačių laiko eilučių pavaizdavimas dvimatės erdvės.
3. Netiesinės prognozavimo lygties gavimo metodai. Tikslumo įvertinimo būdai. Laiko eilučių prognozavimas. Statistiniai modeliai laiko eilučių pasiskirstymams aprašyti. Parametrų kiekio įtaka laiko eilučių pasiskirstymo modelių sudėtingumui charakterizuoti. Stacionarūs ir nestacionarūs atsitiktiniai procesai. Daugiamatės laiko eilutės. Modeliai, naudojami daugiamačių eilučių pasiskirstymams aprašyti, prognozavimo lygtims gauti. Prognozavimo lygties sudėtingumo ir mokymo duomenų kiekio įtaka į prognozavimo tikslumą.
4. Reguliarizuota ir robastrinė regresijos. Optimalių reguliarizacijos ir robastriskumą įvertinančių parametrų radimo būdai. Minimaksinė ir atraminių vektorių regresija. Regresijos lygties sudarymo tipo parinkimas.
5. Prognozavimo lygties tikslumo rodikliai. Lygties sudėtingumo, parametrų įvertinimo būdo ir mokymo duomenų kiekio ryšys. Tikslumo įvertinimo metodai.
6. Klasifikavimo uždavinio statistinis formulavimas. I-os ir II-ros klasifikavimo klaidos ir jų tikimybės, apriorinės klasių tikimybės. Asimptotinė ir bajesinė klasifikavimo klaidos. Neteisingo sprendimo priėmimo kainos ir jų vertinimas sudarant klasifikavimo taisyklės. Klasifikavimas daugelio klasių atveju. Neteisingo klasifikavimo kainų (nuostolių) įvertinimas sudarant klasifikavimo taisyklės.
7. Reguliarizuotos, robastrinės, atraminių vektorių paremtos klasifikavimo taisyklės. Imties tūrio įtaka į klasifikavimo taisyklės tikslumą. Klasifikavimo klaidų rūšys ir jų įvertinimas. Klasifikavimo klaidų įvertinimų tikslumo problema.
8. Vienasluoksnis perceptronas, jo mokymas minimizuojant vidutinę kvadratinę paklaidą ir ryšys su statistiniais prognozavimo ir klasifikavimo metodais. Perceptrono mokymo proceso pagerinimas naudojant statistikoje naudojamas prielaidas apie pasiskirstymo tankio funkcijas.
9. Neparimetriniai vienamačių ir daugiamačių tankių įvertinimo būdai. Duomenų dimensiškumo ir mokymo duomenų tūrio problema. Prognozavimo ir klasifikavimo metodai, paremti neparimetriniais pasiskirstymo tankių įvertinimais.
10. Pagrindinių komponentų metodas. Faktorinė analizė. Duomenų pavaizdavimas mažesnio dimensiškumo erdvėje. Požymių išrinkimas ir išskyrimas. Požymių išrinkimo ir išskyrimo metodai. Požymių išrinkimo ir išskyrimo tikslumo problemos.
11. Daugiamačių pasiskirstymų mišiniai. Klaster analizė. Klaster analizės algoritmai. Problemos, išskylančios atliekant klasterizacijos procedūras. Duomenų apimtys ir dimensiškumo įtaka į klasterizacijos patikimumą. Klasterizacijos kokybės vertinimo kriterijai.
12. Modelio parinkimo problema turint ribotą eksperimentinių duomenų kiekį. Įsivaizduojamoji,

idealoji ir faktinė modelio paklaida regresijos, klasifikacijos ir klasterizacijos uždaviniuose. Jų vertinimas iš empirinių duomenų.

Pagrindinė literatūra

1. L. G. Grimm, P. R. Yarnold. Reading and Understanding Multivariate Statistics. APA publ. Washington, 1994 (paperback, nuo 10 \$, Amazon.com).
2. K. Fukunaga. Introduction to Statistical Pattern Recognition. Academic Press, NY, 2000.
3. Kim H. Esbensen. Multivariate Data Analysis: in Practice : An Introduction to Multivariate Data Analysis and Experimental Design (5th Edition), CAMO Process AS , 2004 (available in internet).
4. S.Raudys. Statistical and Neural Classifiers: An integrated approach to design. Springer, NY, 2001 (galima rasti MIF Informatikos katedros internetiniame puslapyje).
5. S. Raudys, D.Young. Results in statistical discriminant analysis: A review of the former Soviet Union literature. Journal of Multivariate Analysis (ISSN 0047-259X), vol. 89, pp. 1-35, 2004 (galima rasti MIF Informatikos katedros internetiniame puslapyje).
6. J. Kruopis. Matematinė statistika. Mokslas, Vilnius,
7. Š. Raudys, Žinių išgavimas iš duomenų. KU, Klaipėda, 2008 (galima rasti MIF Informatikos katedros internetiniame puslapyje).

Konsultuojančiųjų dėstytojų vardas, pavardė	mokslo laipsnis	pedag. vardas	Svarbiausieji darbai mokslo kryptyje (šakoje) paskelbti per pastaruosius 5 metus
Šarūnas Raudys	Habil. dr.	prof.	S. Raudys, A. Raudys. Pair-wise costs in multi-class perceptrons. <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> (ISSN 0162-8828), vol. 32, pp. 1324-1328, 2010.
			S. Raudys, R. Kybartas, E. K. Zavadskas. Complexity and sample-size issues in multi-category nets of single-layer perceptrons. <i>IEEE Trans. on Neural Networks</i> (ISSN 1045-9227), vol. 21, 784-795, 2010
			S. Raudys. Trainable fusion rules. II. Small sample size case. <i>Neural Networks</i> (ISSN 0893-6080), vol. 19, pp. 1517-1527, 2006. I. Large sample size case. <i>Neural Networks</i> , vol. 19, pp. 1506-1516, 2006.
			Ш.Раудис (S. Raudis), В.Юстицкий (V. Yustitskis). Закон Йеркса—Додсона: связь между стимулированием и успешностью научения. <i>Вопросы психологии (Voprosy Psikhologii)</i> , ISSN 0042-8841), №3, стр. 119-128, 2008.

Patvirtinta Matematikos ir informatikos fakulteto taryboje 2012 m. kovo 6 d., protokolo Nr. 7

Fakulteto tarybos pirmininkas

prof. Gediminas Stepanauskas