GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Statistics II, PG_00021508		
Field of study	Mathematics		
Date of commencement of studies	October 2025	Academic year of realisation of subject	2025/2026
Education level	second-cycle studies	Subject group	Specialty subject group Subject group related to scientific research in the field of study
Mode of study	Full-time studies	Mode of delivery	at the university
Year of study	1	Language of instruction	Polish
Semester of study	2	ECTS credits	5.0
Learning profile	general academic profile	Assessment form	exam
Conducting unit	Department of Nonlinear Analysis and Statistics -> Faculty of Applied Physics and Mathematics -> Faculties of Gdańsk University of Technology		
Name and surname of lecturer (lecturers)	Subject supervisor	dr hab. Karol Dziedziul	
	Teachers	dr hab. Karol Dziedziul	

Lesson types

Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
Number of study hours	30.0	0.0	15.0	0.0	15.0	60

E-learning hours included: 0.0

eNauczanie source addresses:

Moodle ID: 4434 Statystyka II
https://enauczanie.pg.edu.pl/2025/course/view.php?id=4434

Learning activity and number of study hours

Learning activity	Participation in didactic classes included in study plan	Participation in consultation hours	Self-study	SUM
Number of study hours	60	5.0	60.0	125

Subject objectives

The lecture shows the role of the Vapnik Chervonenkis dimension for the correct estimation of regression functions using decision trees, i.e. what makes us have the consistency theorem. Formulas of information criteria are derived: AKAIKE criterion, Fisher criterion. We show the role of the Rao-Cramer inequality in the context of the limits of knowledge, i.e. the role of uncertainty. The laboratory has a completely applied nature. GLM models, regression models using decision trees, neural networks. Random forest concept.

Learning outcomes

Course outcome	Subject outcome	Method of verification
[K7_U06] uses probability distributions and their properties in practical issues, is familiar with the basics of statistics and the basics of statistical data processing	There are many statistical models that can be applied to the same dataset. These models should first be distinguished at the formal level. The student is able to select an appropriate modeling approach, including linear and nonlinear regression, decision trees, and neural networks.	[SU1] Assessment of task fulfilment
[K7_W07] describes well symbolic computation software package and statistical data processing package.	The student is able to assess whether statistical tools are used correctly and understands their formal as well as numerical limitations. The student can identify the conditions under which models are internally consistent and their interpretations are valid.	[SW3] Assessment of knowledge contained in written work and projects
[K7_W01] has enhanced knowledge of basic branches of mathematics,demonstrates knowledge theorem and hypotheses, has understanding of the role and importance of mathematical reasoning structure.	The student is able to perform data diagnostics. This involves a demanding process of statistical inference, including the comparison of information criteria for nested models and related diagnostic procedures.	[SW2] Assessment of knowledge contained in presentation

Subject contents

Course content – lecture

1. Regression estimator, Cantelli Gliwienko's theorems and regression consistency theorems2. Matrix factorization3. Combinatorics according to Vapnik and Czervonenkis. Vapnik Chervonenkis class generation4. Theorem Rao-Blackwell the Hodges-Lehman theorem.5. Kulback Leibler's information, AKAIKE information criteria.6. Cramer Rao's inequality.7. The role of information8. Laboratory - GLM models, regression models using decision trees, neural networks. Random forest concept

Course content – laboratory

The dynamics of the laboratory sessions will be based on classes prepared by Paweł Wieczyński. In addition, students will be required to submit weekly reports reflecting their understanding of key concepts. The laboratory work will focus primarily on examples illustrating definitions, while deeper theoretical connections will be developed during the lectures.

Accordingly, concepts such as deviance will be discussed alongside KullbackLeibler information in the lectures, with maximum likelihood estimation introduced in a largely intuitive manner. The course will then proceed to linear models with LASSO regularisation, logistic regression and generalized linear models, followed by decision trees and neural networks.

Course content – seminar

The seminar will focus on revisiting and consolidating concepts introduced in the lectures and the laboratory sessions.

Prerequisites and co-requisites

Mathematical Statistics and Statistics with SAS, Probabilistics

Assessment methods and criteria

Subject passing criteria	Passing threshold	Percentage of the final grade
Half the exercises half an oral examination	60.0%	100.0%

Recommended reading

Basic literature	J. Bartoszewicz Wykłady ze Statystyki matematycznej PWN Warszawa 1989 Sadanori Konishi, Genshiro Kitagawa: "Information Criteria and Statistical Modeling" Springer Series in Statistics2008 Leo Breiman, Jerome Friedman, Charles J. Stone, R.A. Olshen Classification and Regression Trees Taylor & Francis, 1984
Supplementary literature	R. Zieliński Siedem wykładów wprowadzających do statystyki matematycznej PWN Warszawa 1990 E.L. Lehmann, G. Casella, Theory of Point Estimation Springer Texts inStatistics 2nd Edition 199819
eResources addresses	

Example issues/ example questions/ tasks being completed

1. AKAIKE2. VC dimension, examples, complexity measures (theorem 1.6)3. festimator for decision trees. Its design.4. Cantella Gliwienko's theorem and generalizations from the perspective of the VC dimension.5. PCA or matrix decomposition theorems (applications of PCA)6. State space, statistical space, example

Practical activites within the subject	Not applicable

Document generated electronically. Does not require a seal or signature.