COURSE DESCRIPTION CARD - SYLLABUS

Course name

Advanced Topics in Dimensionality Reduction

Course

Field of study	Year/Semester
Artificial Intelligence	1/2
Area of study (specialization)	Profile of study
	general academic
Level of study	Course offered in
Second-cycle studies	English
Form of study	Requirements
full-time	elective

Number of hours

Lecture	Laboratory classes	Other (e.g. online)
15	15	
Tutorials	Projects/seminars	

Number of credit points

3

Lecturers

Responsible for the course/lecturer:

dr hab. inż. Robert Susmaga

e-mail: robert.susmaga@cs.put.poznan.pl

tel.: 616652934

Faculty of Computing and Telecommunications

Piotrowo 2, 60-965 Poznań

Responsible for the course/lecturer:

Prerequisites

Basic knowledge of linear algebra (simple operations on vectors and matrices) and analytical geometry (graphing simple functions).

Designing, implementing and testing simple computer programs that perform basic vector-matrix operations and generate graphs of basic functions.

(recommended) A fair amount of cognitive curiosity and not less perseverance in pursuing the goals of personal development.

Course objective

1. Providing students with detailed knowledge on:

a) selected aspects of linear algebra, in particular: matrix-vector operations in multidimensional spaces and the eigen-value decompositions of square matrices (EVD) and singular-value decompositions (SVD), along with their applications in the following methods,

b) selected dimensionality reduction methods, including principal component analysis (PCA) and multidimensional scaling (MDS) methods, stochastic embedding methods (SNE and t-SNE) (/t-distributed/ stochastic neighbour embedding').

c) application of data dimension reduction methods to multidimensional data visualization, in particular: to simple visualizations based on barycentric coordinate systems and to visualizations based on advanced dimensional reduction methods.

2. Developing students' skills in:

a) identifying, formulating and solving research problems related to dimensionality reduction and its applications in the visualization of multidimensional data,

b) designing, creating and testing programs implementing the discussed methods.

Course-related learning outcomes

Knowledge

Student:

-- has advanced and in-depth knowledge of methods, tools and programming environments in dimensionality reduction (K2st_W1)

-- has theoretically founded general knowledge related to key issues in the field of data analysis, in particular regarding the analysis of multivariate data, along with their advantages (geometric interpretations) and disadvantages (K2st_W2)

-- has advanced detailed knowledge of selected issues in the field of multivariate data analysis, in particular regarding dimensionality reduction (primarily: in the field of the PCA method), along with their applications in the selection, smoothing and visualization of multidimensional data (primarily: the MDS method, barycentric coordinates) (K2st_W3)

-- knows the development trends and the most important new achievements of computer science, in particular in the field of machine learning and data mining, in which the latest achievements most often use effective optimization algorithms in multidimensional spaces (K2st_W4)

- knows advanced methods, techniques and tools used in solving complex engineering tasks and conducting research in the field of multidimensional data analysis, primarily dimensional reduction and its applications in visualization (K2st_W6).

Skills

Student

-- is able to obtain information from literature, databases and other sources (both in Polish and English), integrate them, interpret and critically evaluate them, draw conclusions and formulate and fully justify opinions (K2st_U1)

-- is able to plan and carry out experiments, including computer measurements and simulations, interpret the obtained results and draw conclusions and formulate and verify hypotheses related to dimensionality reduction problems (K2st_U3)

-- is able to use analytical and simulation-based methods to formulate and solve engineering tasks and simple research problems, in particular regarding the transformation and analysis of multidimensional data (K2st_U4)

-- is able to use and -- when formulating and solving engineering tasks, in particular regarding machine learning and data mining -- integrate knowledge from various areas of mathematics (linear algebra, multidimensional geometry, etc.), also taking into account non-technical aspects (K2st_U5)

-- is able to assess the usefulness of using new achievements (methods and tools) and new IT products, mainly in the fields of analysis and processing of multidimensional data (e.g. methods of reduction / selection of features) (K2st_U6)

-- is able to make a critical analysis of existing technical solutions (in particular, e.g. in the fields of: machine learning and data mining -- solutions requiring effective reduction of data dimensionality) and propose their improvements (improvements) (K2st_U8)

-- is able to assess the usefulness of methods and tools for solving an engineering task related to dimensionality reduction (K2st_U9)

-- is able - using among others conceptually new methods - to solve complex IT tasks related to dimensionality reduction (K2st_U10)

-- is able to -- in accordance with the given specification, taking into account non-technical aspects -- design and implement a complex IT system -- at least in part -- using appropriate methods, techniques and tools, including adapting existing or developing new tools for multidimensional data analysis, in particular: dimensionality reduction (K2st_U11).

Social competences

Student:

-- understands that in computer science, knowledge and skills very quickly become obsolete (K2st_K1),

-- is able to adequately define priorities for the implementation of a task set by her/himself or others (K2st_K2).

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

Formative assessment (laboratory classes): evaluation of the solutions to the assigned programming problems (as they arise).

Final assessment:

-- (laboratory classes): evaluation of the solutions to the assigned programming problems (final),

-- (lectures): evaluation of the results of a written test (45--60 min) with both multiple choice, short answer and (small) computational questions (mostly: micro-problems to be solved in writing).

Programme content

Introduction: The idea of measuring and scaling, basic transformations of variables; multidimensional data and measurements, hidden variables, reduction of variables (dimensionality). The idea of data visualization. Viviani's theorem and barycentric coordinate systems, three-dimensional and four-dimensional. Application of four-dimensional barycentric systems: visualization of confirmation measures and classification accuracy measures. Multidimensional vector spaces, vector norms, dot product of vectors, projection of a vector, angle between vectors, orthogonal vectors. Variable dependence measures, covariance, correlation, similarity measures, distance measures (Euclidean, Minkowski, Mahalanobis, cosine measure). Eigenvalues and their properties, eigenvectors and their properties. The idea of matrix decomposition, eigenvalue decomposition (EVD): construction and basic properties. Matrix singular values and vectors, singular value decomposition (SVD). Concept, procedures and applications of methods: principal component analysis (PCA) and multidimensional scaling (MDS).

Measures and optimization problem (minimization of selected measures) in dimensionality reduction.

The problem of the nearest neighbours, the Kullback-Leibler divergence. Idea, procedures and application of methods: SNE and t-SNE (/ t-distributed / stochastic neighbour embedding).

Teaching methods

Lectures: slide show presentations (theoretical elements, explanations, examples, exercises).

Laboratory classes: designing and creating (in a programming language of one’s choice) programs that solve the assigned problems (which illustrate the ideas and notions presented during the lectures).

Bibliography

Basic

1. J. Koronacki, J. Ćwik: Statystyczne systemy uczące się , WNT, Warszawa, 2005

2. G. Banaszak, W. Gajda: Elementy algebry liniowej część I i II, WNT, Warszawa, 2002

Additional

1. Lecture notes (slide show presentations).

2. K. Sayood (red.): Lossless Compression Handbook, Academic Press, Elsevier Science, San Diego, California, 2003.

3. K. Sayood: Introduction to Data Compression, 3rd Ed., Morgan Kaufmann Publishers, San Francisco, California, 2006.

Breakdown of average student's workload

	Hours	ECTS
Total workload	75	3,0
Classes requiring direct contact with the teacher	30	1,5
Student's own work (literature studies, preparation for laboratory classes, preparation for tests) [1]	45	1,5

[1] delete or add other activities as appropriate