GENERAL

SCHOOL	Engineering			
ACADEMIC UNIT	Computer engineering and informatics			
LEVEL OF STUDIES	Undergraduate (advanced / Integrated MS)			
COURSE CODE	**CEID_**		SEMESTER	Spring
COURSE TITLE	Matrix methods and tools in Data Science			
INDEPENDENT TEACHING ACTIVITIES *if credits are awarded for separate components of the course, e.g. lectures, laboratory exercises, etc. If the credits are awarded for the whole of the course, give the weekly teaching hours and the total credits*			WEEKLY TEACHING HOURS	CREDITS
			4	5
Add rows if necessary. The organisation of teaching and the teaching methods used are described in detail at (d).				
COURSE TYPE *general background, special background, specialised general knowledge, skills development*	special background			
PREREQUISITE COURSES:	The students are expected to have a good understanding of Linear Algebra, Mathematics, Probability & Statistics, Numerical Analysis, Algorithms at the level of the courses in offered in the first 2 years at CEID. Familiarity with the Scientific Computing course (4th year) would also be helpful.			
LANGUAGE OF INSTRUCTION and EXAMINATIONS:	Greek			
IS THE COURSE OFFERED TO ERASMUS STUDENTS	Yes			
COURSE WEBSITE (URL)	(under development) https://eclass.upatras.gr/courses/CEID1164/			

LEARNING OUTCOMES

Learning outcomes *The course learning outcomes, specific knowledge, skills and competences of an appropriate level, which the students will acquire with the successful completion of the course are described.* *Consult Appendix A* • *Description of the level of learning outcomes for each qualifications cycle, according to the Qualifications Framework of the European Higher Education Area* • *Descriptors for Levels 6, 7 & 8 of the European Qualifications Framework for Lifelong Learning and Appendix B* • *Guidelines for writing Learning Outcomes*
The general objective of the course is the presentation of methods and computational tools of linear algebra, emphasizing the solution of problems in Data Science. The mathematical methods target the effective manipulation of the two fundamental objects of the field, namely graphs and matrices. The field is rapidly advancing and has many applications. The course covers advanced matrix methods and their applications in data science. Students will learn about the mathematical foundations of matrices and explore their use in various data science tasks, including data reduction, classification, clustering, and optimization. The course combines theoretical concepts with practical applications, using software tools for hands-on experience. By the end of the course, students would have encountered and used theoretical and practical tools that are essential in the area, they would have a command of these tools' strengths and weaknesses, and they would be able to select methods based on the problem characteristics. They would also be able to apply and combine these techniques and follow the rapidly evolving research literature on the topic.

General Competences	
Taking into consideration the general competences that the degree-holder must acquire (as these appear in the Diploma Supplement and appear below), at which of the following does the course aim?	
Search for, analysis and synthesis of data and information, with the use of the necessary technology *Adapting to new situations* *Decision-making* *Working independently* *Team work* *Working in an international environment* *Working in an interdisciplinary environment* *Production of new research ideas*	*Project planning and management* *Respect for difference and multiculturalism* *Respect for the natural environment* *Showing social, professional and ethical responsibility and sensitivity to gender issues* *Criticism and self-criticism* *Production of free, creative and inductive thinking* *……* *Others…* *…….*
Search for, analysis and synthesis of data and information, with the use of the necessary technology *Adapting to new situations* *Decision-making* *Working in an interdisciplinary environment* *Project planning and management* *Criticism and self-criticism* *Production of free, creative and inductive thinking*	

SYLLABUS

Matrix computations as kernels in Data Science applications. From graphs to vectors to tensors. The many views of matrix multiplication. Classical matrix factorizations. CS and the GSVD. Rank revealing factorizations. Least sqaures and linear regression. Total least squares. Regularization techniques and ridge regression. Solving with iterative methods: descent methods, Krylov subspaces, row projection methods. Dimensionality reduction and clustering applications: Approximating with low rank matrices. Nonnegative least squares and the NMF. Tensors and their decompositions. Randomized numerical linear algebra for very large problems: randomized projections, sketching, CUR, Blendenpik, randomized SVD. Matrix functions and applications in computing centrality indices. Computing the trace and selected matrix elements. The impact of HPC and new architectures: Novel floating point number representations, probabilistic error analyses, stochastic rounding, mixed precision arithmetic. Parallelism in matrix computations. Communication avoiding and asynchronous algorithms. Numerical libraries.

TEACHING and LEARNING METHODS - EVALUATION

DELIVERY *Face-to-face, Distance learning, etc.*	Face-to-face
USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY *Use of ICT in teaching, laboratory education, communication with students*	
TEACHING METHODS *The manner and methods of teaching are described in detail.* *Lectures, seminars, laboratory practice, fieldwork, study and analysis of bibliography, tutorials, placements, clinical practice, art workshop, interactive teaching, educational visits, project, essay writing, artistic creativity, etc.* *The student's study hours for each learning activity are given as well as the hours of non-directed study according to the principles of the ECTS*	see table below

Activity	*Semester workload*
Lectures	39
Tutorials	13
Laboratory practice	13
Independent study	44
Project preparation and writing	13
Examination	3
Course total	***125***

STUDENT PERFORMANCE EVALUATION *Description of the evaluation procedure* *Language of evaluation, methods of evaluation, summative or conclusive, multiple choice questionnaires, short-answer questions, open-ended questions, problem solving, written work, essay/report, oral examination, public presentation, laboratory work, clinical examination of patient, art interpretation, other* *Specifically-defined evaluation criteria are given, and if and where they are accessible to students.*	Evaluation is based on the following factors: 1) Class participation. 2) Written project report including algorithm implementations and numerical experiments. 3) Oral presentation of the report and examination on taught syllabus.

ATTACHED BIBLIOGRAPHY

- G. Strang, Linear Algebra and Learning from Data, Wellesley-Cambridge Press, 2019.
- M. Mahoney, J. Duchi and A.C. Gilbert, eds., The Mathematics of Data, IAS/Park City Mathematics Series, vol. 25, 2018.
- G. Golub and C. Van Loan, "Matrix Computations", 4th edition, Johns Hopkins U, Press, 2013.
- L. Eldén, Matrix Methods in Data Mining and Pattern Recognition, 2nd ed., SIAM, 2019.
- D. Calvettti & E. Somersalo, Mathematics of Data Science: A Computational Approach to Clustering and Classification, SIAM, 2021.
- Phillips, Jeff M., Mathematical Foundations for Data Analysis, Springer Series in the Data Sciences, Springer International Publishing, 2021.
- Peter Grindrod , Mathematical Underpinnings of Analytics, Oxford University Press, 2014.
- A. Kireeva and J.A. Tropp, Randomized matrix computations: themes and variations, California Institute of Technology, 2023.
- E. Gallopoulos, B. Philippe and A. Sameh, Parallelism in Matrix Computations, Springer, 2016.
- A. Laub, Matrix Analysis for Engineers and Scientists, SIAM, 2014.

Selected papers (mostly from) SIAM J. Matrix Analysis, SIAM J. Math. Data Sciences, SIAM J. Scientific Computing, IEEEE TKDE, Data Mining & Knowledge Discovery