

## DAMA502 Module Outline

### 1. GENERAL

<b>SCHOOL</b>	School of Science and Technology		
<b>PROGRAMME</b>	Data Science and Machine Learning		
<b>LEVEL OF STUDIES</b>	Level 7 of the Hellenic and European Qualifications Framework		
<b>MODULE CODE</b>	DAMA502	<b>SEMESTER</b>	1/2/3
<b>MODULE TITLE</b>	Statistics and Optimization		
<b>INDEPENDENT TEACHING ACTIVITIES</b> <i>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</i>		<b>HOURS</b>	<b>CREDITS</b>
Weekly workload: 32-33 hours x 13 weeks		420	15 ECTS
<b>MODULE TYPE</b> <i>Compulsory/Elective/ Mandatory Optional</i>	Compulsory/Elective		
<b>PREREQUISITE MODULES</b>	None		
<b>LANGUAGE OF INSTRUCTION and EXAMINATIONS</b>	English		
<b>IS THE MODULE OFFERED TO ERASMUS STUDENTS</b>	Yes		
<b>MODULE WEBSITE (URL)</b>	The Module has a dedicated space in HOU's digital learning platform ( <a href="http://courses.eap.gr">http://courses.eap.gr</a> , <a href="http://study.eap.gr">http://study.eap.gr</a> ), which students and tutors can access using their credentials.		

### 2. LEARNING OUTCOMES

<p><b>Learning outcomes</b>  <i>The learning outcomes, specific knowledge, skills and competences of an appropriate level, which the students will acquire with the successful completion of the Module are described.</i>  <i>Consult Appendix A</i></p> <ul style="list-style-type: none"> <li>• <i>Description of the level of learning outcomes for each qualifications cycle, according to the Qualifications Framework of the European Higher Education Area</i></li> <li>• <i>Descriptors for Levels 6, 7 &amp; 8 of the European Qualifications Framework for Lifelong Learning and Appendix B</i></li> <li>• <i>Guidelines for writing Learning Outcomes</i></li> </ul>
<p><b>Knowledge:</b>          Upon successful completion of the Module, students will be able to:</p> <ul style="list-style-type: none"> <li>- Recognize that basic mathematical pillars for machine learning are probability theory, statistics and optimization and apply analytical and computational tools.</li> <li>- Summarize the properties of single variate and multivariate Gaussian distribution, find marginals and conditionals as well as transformations of the Gaussian function.</li> <li>- Focus on the binomial Bernoulli distribution and detail the Beta distribution.</li> <li>- Summarize the conjugate priors connected through Bayes theorem.</li> <li>- Explain what sufficient statistics is and outline the exponential family of distributions.</li> <li>- Perform a change of random variables and find the new distribution function.</li> <li>- List basic statistical analysis techniques</li> <li>- Perform hypothesis testing</li> <li>- Identify outliers</li> <li>- Recall how to find minima of a function of a single variable.</li> <li>- Summarize the procedure to find the minimum of a multivariate function using the gradient descent algorithm</li> <li>- Explain how to perform stochastic gradient descent and what are its advantages and limitations</li> </ul>

- compared to the gradient descent method
- Describe what are the Lagrange multipliers and explain how they are used in constrained optimization.
- Describe convex optimization
- Use SageMath to find the minimum of a multivariate function and t to minimize a function with constraints. Evaluate the performance of a model
- Acquire knowledge on how to extract useful information from data visualization.
- Understand how to visualize 1D and 2D data using binning, density plots, scatter plots, and box plots.

**Skills:**

Upon successful completion of the Module, students will be able to:

- Perform change of variables in probability distributions and derive the resulting distribution functions.
- Conduct hypothesis testing using standard statistical procedures.
- Detect and interpret outliers in datasets.
- Use SageMath to minimize multivariate functions and solve constrained optimization problems.
- Evaluate the performance of machine learning models using optimization and statistical metrics.
- Generate appropriate visualizations (e.g., histograms, density plots, scatter plots) for 1D and 2D data using appropriate tools.

**Competences:**

Upon successful completion of the Module, students will be able to:

- Select and apply appropriate probabilistic, statistical, and optimization methods to analyze and solve problems in machine learning.
- Combine statistical reasoning with computational tools to interpret real-world data and make informed decisions.
- Critically assess the suitability of optimization algorithms (e.g., SGD vs. GD) for training specific machine learning models.
- Solve practical optimization problems using constrained methods (e.g., Lagrange multipliers) within a machine learning context.
- Interpret and communicate findings from data visualizations to support data-driven conclusions.
- Work autonomously with mathematical software tools (e.g., SageMath) to analyze data and validate learning algorithms.

**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 Module aim?*

<i>Search for, analysis and synthesis of data and information, with the use of the necessary technology</i>	<i>Project planning and management</i>
<i>Adapting to new situations</i>	<i>Respect for difference and multiculturalism</i>
<i>Decision-making</i>	<i>Respect for the natural environment</i>
<i>Working independently</i>	<i>Showing social, professional and ethical responsibility and sensitivity to gender issues</i>
<i>Team work</i>	<i>Criticism and self-criticism</i>
<i>Working in an international environment</i>	<i>Production of free, creative and inductive thinking</i>
<i>Working in an interdisciplinary environment</i>	
<i>Production of new research ideas</i>	

**The general skills that the students will acquire are:**

- 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
- Project planning and management
- Showing social, professional and ethical responsibility and sensitivity to gender issues
- Production of free, creative and inductive thinking

**3. SYLLABUS**

**Purpose of Module**

The students will learn the basic mathematical tools necessary for Machine Learning (ML). These include basic concepts from probability theory, introductory statistics and convex optimization. Also, the student will learn basic visualization techniques of 1D and 2D data. Overall, a student without prior knowledge of these mathematical areas will be able to form a background to understand ML techniques while students with prior mathematical knowledge will be able to go much deeper in application of mathematics in ML. The mathematical study will be supplemented by computational software that will enable both analytical and numerical evaluations.

The key subjects of the module are "Probability Theory and Statistics", "Convex Optimization" and "Visualization".

**4. TEACHING and LEARNING METHODS - EVALUATION**

<p><b>DELIVERY</b> <i>Face-to-face, Distance learning, etc.</i></p>	<ul style="list-style-type: none"> <li>- Distance teaching and learning with three (3) Group Counseling Meetings (GCMs) of 4-hour duration during the academic semester on weekends.</li> <li>- Personal communication and feedback (advisory role of Adjunct Faculty).</li> </ul>													
<p><b>USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY</b> <i>Use of ICT in teaching, laboratory education, communication with students</i></p>	<p><u>During GCMs and teaching the following are used:</u></p> <ul style="list-style-type: none"> <li>- Remote meetings tools (webex, Teams),</li> <li>- Presentation software (PowerPoint, educational video - animations etc.),</li> <li>- Specialized software/databases for the subjects under study.</li> </ul> <p>In addition, students use office automation tools, web browsers and e-reader for digital books.</p> <p><u>Communication with the students is supported by:</u></p> <ul style="list-style-type: none"> <li>- The digital platform of HOU (<a href="https://courses.eap.gr/login/index.php">https://courses.eap.gr/login/index.php</a> / <a href="https://study.eap.gr/login/index.php">https://study.eap.gr/login/index.php</a>) (course information, educational material posts, announcements, messages, examination results, user groups, discussion forums etc.).</li> <li>- e-mail and messaging.</li> </ul>													
<p><b>TEACHING METHODS</b> <i>The manner and methods of teaching are described in detail.</i> <i>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.</i> <i>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</i></p>	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="width: 70%;">Activity</th> <th style="width: 30%;">Semester Workload</th> </tr> </thead> <tbody> <tr> <td>3 GCMs (x 4 hours)</td> <td>12</td> </tr> <tr> <td>3 Semester Assignments (x 27 hours)</td> <td>81</td> </tr> <tr> <td>Individual Study time (25 hours x 13 weeks)</td> <td>325</td> </tr> <tr> <td>Final examination</td> <td>3</td> </tr> <tr> <td><b>Total Workload</b></td> <td><b>421</b></td> </tr> </tbody> </table>		Activity	Semester Workload	3 GCMs (x 4 hours)	12	3 Semester Assignments (x 27 hours)	81	Individual Study time (25 hours x 13 weeks)	325	Final examination	3	<b>Total Workload</b>	<b>421</b>
Activity	Semester Workload													
3 GCMs (x 4 hours)	12													
3 Semester Assignments (x 27 hours)	81													
Individual Study time (25 hours x 13 weeks)	325													
Final examination	3													
<b>Total Workload</b>	<b>421</b>													
<p><b>STUDENT PERFORMANCE EVALUATION</b> <i>Description of the evaluation procedure</i> <i>Language of evaluation, methods of evaluation, summative or conclusive, multiple choice questionnaires, short-answer questions, open-ended</i></p>	<p><b>Students' evaluation – Grade assessment of a Module:</b></p> <p>a. Three (3) Semester Assignments (A) which contribute equally to the final grade with a value of 10% each, for a total of 30%.</p> <p>The scoring of assignments is activated only if the student succeeds in an overall score equal to or above the base (<math>\geq 50\%</math>) in the final or repeat exams.</p> <p>b. Final or repeat exams (E) contribute to the final grade of the module by 70%.</p>													

*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.*

The Final Grade of the module is calculated as follows (with 10 being the maximum Grade):

$$\text{Final Grade} = (A1 \times 10\%) + (A2 \times 10\%) + (A3 \times 10\%) + (E \times 70\%)$$

**Language of evaluation:** English

## 5. INDICATIVE BIBLIOGRAPHY

- *Recommended bibliography:*

- Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong. Mathematics for Machine Learning. Cambridge University Press, 2020
- Paul Zimmermann, Alexandre Casamayou, Nathann Cohen, Guillaume Connan, Thierry Dumont, Laurent Fousse, François Maltey, Matthias Meulien, Marc Mezzarobba, Clément Pernet, Nicolas M. Thiéry, Erik Bray, John Cremona, Marcelo Forets, Alexandru Ghitza, Hugh Thomas. Computational Mathematics with SageMath. SIAM, 2018
- Michael R. Berthold, Christian Borgelt, Frank Höppner and Frank Klawonn. (2010). Guide to Intelligent Data Analysis. How to Intelligently Make Sense of Real Data. Springer.
- EMC, E. S., & EMC, E. S. (Eds.). (2015). Data science and big data analytics: Discovering, analyzing, visualizing and presenting data. Wiley.

Additional digital (and multimedia) material will be made available online.