295102 - 295II012 - Data Analysis & Pattern Recognition

Coordinating unit: 295 - EEBE - Barcelona East School of Engineering
Teaching unit: 723 - CS - Department of Computer Science
707 - ESAII - Department of Automatic Control
749 - MAT - Department of Mathematics

Academic year: 2019
Degree: MASTER'S DEGREE IN CHEMICAL ENGINEERING (Syllabus 2019). (Teaching unit Compulsory)
MASTER'S DEGREE IN INTERDISCIPLINARY AND INNOVATIVE ENGINEERING (Syllabus 2019).
(Teaching unit Compulsory)

ECTS credits: 6  Teaching languages: English

Teaching staff
Coordinator: Raúl Benítez Iglesias
Others: Primer quadrimestre:
RAUL BENITEZ IGLESIAS - T11, T12
GERARD ESCUDERO BAKX - T11, T12
SAMIR KANAAN IZQUIERDO - T11, T12
FRANCESC POZO MONTERO - T11, T12

Degree competences to which the subject contributes

Specific:
CEMUEQ-01. To apply knowledge of mathematics, physics, chemistry, biology and other natural sciences, obtained through study, experience and practice, with critical reasoning, to establish economically viable solutions to technical problems

CEMUEQ-03. Conceptualize engineering models, apply innovative methods in the resolution of problems and adequate computer applications, for the design, simulation, optimization and control of processes and systems

CEMUEII-02. Apply techniques of pattern recognition, artificial intelligence and statistical data analysis that allow decisions to be made objectively, quantitatively and reproducibly in problems of a multidisciplinary nature.

Generical:
CGMUEQ-04. To carry out the appropriate research, undertake the design and manage the development of engineering solutions, in new or little known environments, relating creativity, originality, innovation and technology transfer

CGMUEQ-05. Know how to establish mathematical models and develop them through appropriate information technology, as a scientific and technological base for the design of new products, processes, systems and services, and for the optimization of others already developed

CGMUEII-01. Participate in technological innovation projects in multidisciplinary problems, applying mathematical, analytical, scientific, instrumental, technological and management knowledge.

Transversal:
05 TEQ. TEAMWORK. Being able to work as a team player, either as a member or as a leader. Contributing to projects pragmatically and responsibly, by reaching commitments in accordance to the resources that are available.

06 URI. EFFECTIVE USE OF INFORMATION RESOURCES. Managing the acquisition, structure, analysis and display of information from the own field of specialization. Taking a critical stance with regard to the results obtained.

03 TLG. THIRD LANGUAGE. Learning a third language, preferably English, to a degree of oral and written fluency that fits in with the future needs of the graduates of each course.
295102 - 295II012 - Data Analysis & Pattern Recognition

Teaching methodology

The methodology of the course combines theory lessons, laboratory sessions and autonomous learning through the development of projects and the analysis of real applications.

Learning objectives of the subject

.

Study load

<table>
<thead>
<tr>
<th>Total learning time: 150h</th>
<th>Hours large group: 34h</th>
<th>22.67%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours medium group:</td>
<td>0h</td>
<td>0.00%</td>
</tr>
<tr>
<td>Hours small group:</td>
<td>20h</td>
<td>13.33%</td>
</tr>
<tr>
<td>Guided activities:</td>
<td>0h</td>
<td>0.00%</td>
</tr>
<tr>
<td>Self study:</td>
<td>96h</td>
<td>64.00%</td>
</tr>
</tbody>
</table>
## Content

<table>
<thead>
<tr>
<th>Exploratory data analysis</th>
<th>Learning time: 10h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 6h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 4h</td>
</tr>
</tbody>
</table>

**Description:**
- Data visualization (histograms, box-plot, qq-plot, multi-dimensional scatter plots, etc)
- Data clustering (k-means, agglomerative, Gaussian Mixture Models)
- Dimensionality reduction and Principal Component Analysis
- Data representation and feature extraction
- Data metrics, distances, norms, etc.

**Related activities:**
- Laboratory session 1: Data visualization and cluster analysis
- Laboratory session 2: Principal Component Analysis and dimensionality reduction

<table>
<thead>
<tr>
<th>Supervised machine learning</th>
<th>Learning time: 12h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 8h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 4h</td>
</tr>
</tbody>
</table>

**Description:**
- Introduction to Classification/Regression problems
- Distance-based methods: kNN & Centroids algorithm
- Probabilistic methods: Naïve Bayes & LDA
- Rule-based methods: Decision Trees & AdaBoost
- Hyperplane-based methods: kernels & SVM

**Related activities:**
- Laboratory session 3: Supervised classification I
- Laboratory session 4: Supervised classification II
<table>
<thead>
<tr>
<th>Performance evaluation</th>
<th>Learning time: 12h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 8h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 4h</td>
</tr>
</tbody>
</table>

**Description:**
- Type I and type II errors
- Univariate and multivariate hypothesis testing approaches
- Statistical inference and parameter estimation (Maximum-likelihood, Bayesian, bootstrapping)
- Validation procedures: cross-validation; leave-one-out, etc.

**Related activities:**
- Laboratory session 5: Hypothesis testing
- Laboratory session 6: Model validation procedures

**Specific objectives:**

<table>
<thead>
<tr>
<th>Neural networks and deep learning</th>
<th>Learning time: 10h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 6h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 4h</td>
</tr>
</tbody>
</table>

**Description:**
- Introduction to artificial neural networks (ANNs) and deep learning
- Feed-forward ANNs for classification and regression
- Training ANNs: backpropagation algorithm, optimization stages, advanced strategies (network complexity, early stopping, dropout, weight regularization)

**Related activities:**
- Laboratory session 7: Artificial Neural Networks
- Laboratory session 8: Deep Learning

<table>
<thead>
<tr>
<th>Advanced topics and applications</th>
<th>Learning time: 8h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 4h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 4h</td>
</tr>
</tbody>
</table>

**Description:**
Seminars by experts, application projects, analysis of publications, news and trending topics.
Qualification system

Partial exam 30%
Final exam 30%
Projects and exercises 40%

"Those students who meet the requirements set by the EEKE in their Assessment and Permanence Regulations will be able to access the re-assessment test (https://eebe.upc.edu/ca/estudis/estudis-de-master/documents-masters/assessment-and-academic-progress-regulations-for-bachelors-and-masters-degrees-at-the-eebe.pdf)"

Bibliography

Basic:

Others resources:
Python programming: https://www.python.org/
Numpy mathematical libraries: http://www.numpy.org/
Graphical representation: https://matplotlib.org/
Data repository: https://archive.ics.uci.edu/ml/index