295102 - 295II012 - Data Analysis & Pattern Recognition

Coordinating unit: 295 - EEBE - Barcelona East School of Engineering
Teaching unit: 707 - ESAII - Department of Automatic Control
Academic year: 2018
Degree: 
ECTS credits: 6  
Teaching languages: English

Teaching staff
Coordinator: Raúl Benítez Iglesias
Others: Gerard Escudero Bakx
         Samir Kanaan Izquierdo
         Francesc Pozo Montero

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></td>
<td>Hours medium group: 0h</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Hours small group: 20h</td>
<td>13.33%</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 0h</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Self study: 96h</td>
<td>64.00%</td>
</tr>
</tbody>
</table>
# Exploratory data analysis

**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

## Learning time:
- Theory classes: 6h
- Laboratory classes: 4h

---

# Supervised machine learning

**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

## Learning time:
- Theory classes: 8h
- Laboratory classes: 4h
### Performance evaluation

<table>
<thead>
<tr>
<th>Description:</th>
</tr>
</thead>
</table>
| • 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. |

<table>
<thead>
<tr>
<th>Related activities:</th>
</tr>
</thead>
</table>
| Laboratory session 5: Hypothesis testing  
Laboratory session 6: Model validation procedures |

<table>
<thead>
<tr>
<th>Specific objectives:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Learning time: 12h</th>
</tr>
</thead>
</table>
| Theory classes: 8h  
Laboratory classes: 4h |

### Neural networks and deep learning

<table>
<thead>
<tr>
<th>Description:</th>
</tr>
</thead>
</table>
| • 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)  
• Specialized architectures: recurrent neural networks, autoencoders, generative adversarial networks, convolutional neural networks |

<table>
<thead>
<tr>
<th>Related activities:</th>
</tr>
</thead>
</table>
| Laboratory session 7: Artificial Neural Networks  
Laboratory session 8: Deep Learning |

<table>
<thead>
<tr>
<th>Learning time: 10h</th>
</tr>
</thead>
</table>
| Theory classes: 6h  
Laboratory classes: 4h |

### Advanced topics and applications

<table>
<thead>
<tr>
<th>Description:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seminars by experts, application projects, analysis of publications, news and trending topics.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Learning time: 8h</th>
</tr>
</thead>
</table>
| Theory classes: 4h  
Laboratory classes: 4h |
Qualification system

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

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