270717 - IDADM - Intelligent Data Analysis and Data Mining

Coordinating unit: 270 - FIB - Barcelona School of Informatics
Teaching unit: 723 - CS - Department of Computer Science
Academic year: 2019
Degree: MASTER'S DEGREE IN ARTIFICIAL INTELLIGENCE (Syllabus 2012). (Teaching unit Optional)
MASTER'S DEGREE IN ARTIFICIAL INTELLIGENCE (Syllabus 2009). (Teaching unit Optional)
MASTER'S DEGREE IN ARTIFICIAL INTELLIGENCE (Syllabus 2009). (Teaching unit Optional)
MASTER'S DEGREE IN ARTIFICIAL INTELLIGENCE (Syllabus 2017). (Teaching unit Optional)
ECTS credits: 4.5

Prior skills

Students are expected to have at least some basic background in the area of artificial intelligence and, more specifically, with the areas of Machine Learning and Computational Intelligence.
Some basic knowledge of probability theory and statistics would be beneficial.
Other than this, the course is open to students and researchers of all types of background

Teaching methodology

This course will build on different teaching methodology (TM) aspects, including:
TM1: Expositive seminars
TM2: Expositive-participative seminars
TM3: Orientation for individual assignments (essays)
TM4: Individual tutorization

Learning objectives of the subject

1. Presenting DM as a process that should involve a methodology id applied at its best.
2. Introducing the students to the new concept of DM for processes, called Process Mining.
3. Delving into some detail in one of the stages of DM: data exploration.
4. Dealing in detail with the problem of data visualization for exploration as a key issue in DM.
5. Introducing the students to the basics of probability theory as applied in Intelligent Data Analysis (IDA)
6. Introducing the students to the probabilistic variant of IDA in the form of Statistical Machine Learning, both for supervised and unsupervised learning models.
7. Dealing in detail with different unsupervised models for data visualization, including case studies.
9. Approaching the multi-faceted concept of data mining (DM) from different perspectives.

Study load

<table>
<thead>
<tr>
<th>Total learning time: 112h 30m</th>
<th>Hours large group:</th>
<th>15h</th>
<th>13.33%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hours medium group:</td>
<td>15h</td>
<td>13.33%</td>
</tr>
<tr>
<td></td>
<td>Hours small group:</td>
<td>7h 30m</td>
<td>6.67%</td>
</tr>
<tr>
<td></td>
<td>Guided activities:</td>
<td>3h</td>
<td>2.67%</td>
</tr>
<tr>
<td></td>
<td>Self study:</td>
<td>72h</td>
<td>64.00%</td>
</tr>
</tbody>
</table>
## Content

### Introduction to the concept of data mining (DM).

**Degree competences to which the content contributes:**

**Description:**
DM is a multi-faceted concept that requires discussion and clarification. We will do this at the beginning of the course.

### DM as a methodology.

**Degree competences to which the content contributes:**

**Description:**
We argue that DM should not be focused on the concept of data analysis/modeling, but, instead, should be treated as a methodology with diverse inter-related stages.

### DM for processes: Process Mining.

**Degree competences to which the content contributes:**

**Description:**
A new development in DM methodologies is that which deals with one specifically suited for processes. It is called Process Mining and will be described and discussed in this course.

### Data exploration in DM.

**Degree competences to which the content contributes:**

**Description:**
One of the main stages of well-structures DM methodologies is Data exploration. It will be discussed as a preamble to data visualization.

### Basics of probability theory in Intelligent Data Analysis (IDA)

**Degree competences to which the content contributes:**

**Description:**
For a long time in the last half-century, multivariate statistics and artificial intelligence (mostly in the field of machine learning) have developed in parallel without fully meeting. Statistical machine learning has bridged that field over the last two decades. We introduce it by first providing some basic principles of probability theory (Bayesian inference).

### Data visualization for exploration.

**Degree competences to which the content contributes:**
One of the aspects of the problem of data exploration is data visualization. It has a research 'life' of its own as it involves not only computer-based mathematical models, but also natural perception and processing.

### Statistical Machine Learning for IDA: supervised models.

**Degree competences to which the content contributes:**

**Description:**
Once the basics of Bayesian inference are set, we will delve into the field of Statistical Machine Learning for IDA, starting with supervised learning models, with an emphasis on feed-forward artificial neural networks.

### Statistical Machine Learning for IDA: unsupervised models.

**Degree competences to which the content contributes:**

**Description:**
Once the basics of Bayesian inference and of Statistical Machine Learning for IDA in supervised models are set, we will continue with unsupervised models, focusing on self-organizing maps and related models.

### Unsupervised models for data visualization, with case studies.

**Degree competences to which the content contributes:**

**Description:**
In the final item of the contents of the course, we will bring statistical machine learning and data visualization together by discussing some probabilistic unsupervised learning models for data visualization, including some case studies as an example.
## Planning of activities

<table>
<thead>
<tr>
<th>Activity</th>
<th>Hours</th>
<th>Description</th>
<th>Specific objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction to Data Mining and its Methodologies</strong></td>
<td>15h</td>
<td>Introduction to Data Mining as a general concept and to its methodologies for practical implementation</td>
<td>1, 9</td>
</tr>
<tr>
<td><strong>Process Mining</strong></td>
<td>9h</td>
<td>Introduction to the novel concept of Process Mining and its application within the DM framework.</td>
<td>2</td>
</tr>
<tr>
<td><strong>Data Visualization</strong></td>
<td>15h</td>
<td>As part of the DM stage of Data Exploration, we focus in the problem of Data Visualization.</td>
<td>3, 4</td>
</tr>
<tr>
<td><strong>Basics of probability theory for intelligent data analysis</strong></td>
<td>15h</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Theory classes: 6h</th>
<th>Practical classes: 0h</th>
<th>Laboratory classes: 0h</th>
<th>Guided activities: 0h</th>
<th>Self study: 9h</th>
<th>Theory classes: 3h</th>
<th>Practical classes: 0h</th>
<th>Laboratory classes: 0h</th>
<th>Guided activities: 0h</th>
<th>Self study: 6h</th>
<th>Theory classes: 6h</th>
<th>Practical classes: 0h</th>
<th>Laboratory classes: 0h</th>
<th>Guided activities: 0h</th>
<th>Self study: 9h</th>
</tr>
</thead>
</table>
Description:
Introduction to probability theory for intelligent data analysis, with a focus on Bayesian statistics

Specific objectives:
5

Statistical Machine Learning methods

<table>
<thead>
<tr>
<th>Hours: 28h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 12h</td>
</tr>
<tr>
<td>Practical classes: 0h</td>
</tr>
<tr>
<td>Laboratory classes: 0h</td>
</tr>
<tr>
<td>Guided activities: 0h</td>
</tr>
<tr>
<td>Self study: 16h</td>
</tr>
</tbody>
</table>

Description:
The meeting of statistics and machine learning: Statistical Machine Learning methods, from the point of view of both supervised and supervised learning

Specific objectives:
5

SML in data visualization, with case studies

<table>
<thead>
<tr>
<th>Hours: 15h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 6h</td>
</tr>
<tr>
<td>Practical classes: 0h</td>
</tr>
<tr>
<td>Laboratory classes: 0h</td>
</tr>
<tr>
<td>Guided activities: 0h</td>
</tr>
<tr>
<td>Self study: 9h</td>
</tr>
</tbody>
</table>

Description:
We merge the topics of SML and data visualization, illustrating its use with some real case studies

Specific objectives:
7

Qualification system

The course will be evaluated through a final essay that will take one of these three modalities:
1. State of the art on an specific IDA-DM topic
2. Evaluation of an IDA-DM software tool with original experiments
3. Pure research essay, with original experimental content
Bibliography

Basic:


Complementary:

