270658 - KMLMM - Kernel-Based Machine Learning and Multivariate Modelling

Coordinating unit: 270 - FIB - Barcelona School of Informatics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research
723 - CS - Department of Computer Science

Academic year: 2019
Degree: MASTER’S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Teaching unit Optional)
ECTS credits: 6
Teaching languages: English

Teaching methodology
Learning is done through a combination of theoretical explanations and their application to practising exercises and real cases. The lectures will develop the necessary scientific knowledge, including its application to problem solving. These problems constitute the practical work of the students on the subject, which will be developed as autonomous learning. The software used will be primarily R.

Learning objectives of the subject
1. Understand the foundations of Kernel-Based Learning Methods
2. Get acquainted with specific kernel-based methods, such as the Support Vector Machine
3. Know methods for kernelizing existing statistical or machine learning algorithms
4. Know the theoretical foundations of kernel functions and kernel methods
5. Understanding the foundations of the Multivariate Modeling
6. Get acquainted with multivariate modeling from latent components methods
7. Know modeling techniques for broad data matrices (p>n)

Study load

<table>
<thead>
<tr>
<th>Total learning time: 150h</th>
<th>Hours large group: 24h</th>
<th>16.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours medium group:</td>
<td>12h</td>
<td>8.00%</td>
</tr>
<tr>
<td>Hours small group:</td>
<td>12h</td>
<td>8.00%</td>
</tr>
<tr>
<td>Guided activities:</td>
<td>6h</td>
<td>4.00%</td>
</tr>
<tr>
<td>Self study:</td>
<td>96h</td>
<td>64.00%</td>
</tr>
</tbody>
</table>
Introduction to Kernel-Based Learning

Degree competences to which the content contributes:
Description:
This topic introduces the student the foundations of Kernel-Based Learning focusing on Kernel Linear Regression.

The Support Vector Machine (SVM)

Degree competences to which the content contributes:
Description:
This topic develops Support Vector Machine (SVM) for classification, regression and novelty detection.

Kernels: properties & design

Degree competences to which the content contributes:
Description:
This topic defines kernel functions, their properties and construction. Introduces specific kernels for different data types, such as real vectors, categorical information, feature subsets, strings, probability distributions and graphs.

Kernelizing ML algorithms

Degree competences to which the content contributes:
Description:
This topic reviews different techniques for kernelizing existent algorithms.

Theoretical underpinnings

Degree competences to which the content contributes:
Description:
This topic reviews the basic theoretical underpinnings of kernel-based methods, focusing on statistical learning theory.

Introduction to Multivariate Modeling: Principal Component Regression

Degree competences to which the content contributes:
Description:
Extending the Regression to the multivariate case. Principal Component Analysis and Regression. The NIPALS algorithm.
Canonical Correlation Analysis

Degree competences to which the content contributes:
Description:
Canonical Correlation Analysis and the Interbatteries Analysis

Partial Least Squares Regression

Degree competences to which the content contributes:
Description:
Modeling Univariate Partial Least Squares Regression. Algorithms and properties

Multivariate Partial Least Squares Regression

Degree competences to which the content contributes:
Description:
Modeling Multivariate Partial Least Squares Regression. Algorithms and properties

Modeling categorical data with PLSR

Degree competences to which the content contributes:
Description:
Modeling categorical data with PLSR

Partial Least Squares Path Modeling

Degree competences to which the content contributes:
Description:
Introduction to path models and Partial Least Squares Path Modeling
## Planning of activities

<table>
<thead>
<tr>
<th>Topic</th>
<th>Hours</th>
<th>Theory classes</th>
<th>Practical classes</th>
<th>Laboratory classes</th>
<th>Guided activities</th>
<th>Self study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to Kernel-Based Learning</td>
<td>10h</td>
<td>4h</td>
<td>0h</td>
<td>0h</td>
<td>0h</td>
<td>6h</td>
</tr>
<tr>
<td><strong>Specific objectives:</strong></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The SVM for classification, regression and</td>
<td>9h</td>
<td>3h</td>
<td>0h</td>
<td>0h</td>
<td>0h</td>
<td>6h</td>
</tr>
<tr>
<td>novelty detection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Specific objectives:</strong></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kernels: properties &amp; design</td>
<td>10h</td>
<td>4h</td>
<td>0h</td>
<td>0h</td>
<td>0h</td>
<td>6h</td>
</tr>
<tr>
<td><strong>Specific objectives:</strong></td>
<td>1, 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice class (I): the SVM</td>
<td>9h</td>
<td>3h</td>
<td>0h</td>
<td>0h</td>
<td>0h</td>
<td>6h</td>
</tr>
<tr>
<td><strong>Specific objectives:</strong></td>
<td>1, 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Kernelizing ML algorithms

**Specific objectives:**
3, 4

**Hours:** 10h  
- Theory classes: 4h  
- Practical classes: 0h  
- Laboratory classes: 0h  
- Guided activities: 0h  
- Self study: 6h

---

# Practice class (II): kernel design & other KBL methods

**Specific objectives:**
3, 4

**Hours:** 9h  
- Theory classes: 3h  
- Practical classes: 0h  
- Laboratory classes: 0h  
- Guided activities: 0h  
- Self study: 6h

---

# Theoretical underpinnings

**Specific objectives:**
1, 4

**Hours:** 10h  
- Theory classes: 4h  
- Practical classes: 0h  
- Laboratory classes: 0h  
- Guided activities: 0h  
- Self study: 6h

---

# Introduction to Multivariate Modeling. Principal Component Regression.

**Hours:** 9h  
- Theory classes: 3h  
- Practical classes: 0h  
- Laboratory classes: 0h  
- Guided activities: 0h  
- Self study: 6h

---

# Canonical Correlation Analysis

**Hours:** 10h  
- Theory classes: 4h  
- Practical classes: 0h  
- Laboratory classes: 0h  
- Guided activities: 0h  
- Self study: 6h
## Partial Least Squares Regression 1.

| Hours | Theory classes: 3h | Practical classes: 0h | Laboratory classes: 0h | Guided activities: 0h | Self study: 6h |

## Partial Least Squares Regression 2.

| Hours | Theory classes: 4h | Practical classes: 0h | Laboratory classes: 0h | Guided activities: 0h | Self study: 6h |

## Modeling categorical data with PLSR.

| Hours | Theory classes: 3h | Practical classes: 0h | Laboratory classes: 0h | Guided activities: 0h | Self study: 6h |

## Partial Least Squares Path Modeling.

| Hours | Theory classes: 4h | Practical classes: 0h | Laboratory classes: 0h | Guided activities: 0h | Self study: 6h |

## Modeling complex data with latent variables.

| Hours | Theory classes: 2h | Practical classes: 0h | Laboratory classes: 0h | Guided activities: 0h 12m | Self study: 6h |

## Evaluation quiz

| Hours | Guided activities: 3h | Self study: 12h |
Specific objectives:
1, 2, 3, 4, 5, 6, 7

Qualification system

The course evaluation will be based on the marks obtained in the practices during the year plus the marks obtained in the written test for global evaluation.

Each practice will lead to the drafting of the corresponding written report which will be evaluated by the teachers, resulting in a mark denoted P.

The written test will be the last day of class and will evaluate the assimilation of the basic concepts on the subject, resulting in a mark denoted T.

The final mark will be obtained as:

60% x P + 40% x T
## Bibliography

### Basic:


### Complementary:


### Others resources:

**Hyperlink**

- [http://cran.r-project.org/](http://cran.r-project.org/)