270659 - OTDM - Optimization Techniques for Data Mining

Coordinating unit: 270 - FIB - Barcelona School of Informatics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research
Academic year: 2018
Degree: MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Teaching unit Optional)
ECTS credits: 6
Teaching languages: Catalan

Degree competences to which the subject contributes

Basic:
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.

Specific:
CEC1. Ability to apply scientific methodologies in the study and analysis of phenomena and systems in any field of Information Technology as well as in the conception, design and implementation of innovative and original computing solutions.
CEC2. Capacity for mathematical modelling, calculation and experimental design in engineering technology centres and business, particularly in research and innovation in all areas of Computer Science.
CEE3.2. Capability to use a wide and varied spectrum of algorithmic resources to solve high difficulty algorithmic problems.
CEE3.3. Capability to understand the computational requirements of problems from non-informatics disciplines and to make significant contributions in multidisciplinary teams that use computing.
CEE5.3. Capability to work in interdisciplinary engineering services teams and, provided the necessary domain experience, capability to work autonomously in specific service systems.

General:
CG1. Capability to apply the scientific method to study and analyse of phenomena and systems in any area of Computer Science, and in the conception, design and implementation of innovative and original solutions.
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.

Transversal:
CTR2. SUSTAINABILITY AND SOCIAL COMMITMENT : Capability to know and understand the complexity of the typical economic and social phenomena of the welfare society. Capacity for being able to analyze and assess the social and environmental impact.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Teaching methodology

The students will have available all the course material.

About two thirds of lecture time will be devoted to optimization algorithms and their properties, and the rest will be for presenting and solving exercises and problems.

Lab sessions (using AMPL) will be devoted to the solution of some data mining applications.

Learning objectives of the subject
270659 - OTDM - Optimization Techniques for Data Mining

1. Discerning what is an optimization problem and its type and having a basic knowledge of optimization algorithms
2. Formulating optimization problems and representing them through a modeling language
3. Choosing an adequate solver type for a given problem
4. Using publicly available and commercial solvers. Interpreting results from solvers and communicating in writing results from optimization

### Study load

<table>
<thead>
<tr>
<th>Total learning time: 150h</th>
<th>Theory classes:</th>
<th>24h</th>
<th>16.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practical classes:</td>
<td>12h</td>
<td>8.00%</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes:</td>
<td>12h</td>
<td>8.00%</td>
</tr>
<tr>
<td></td>
<td>Guided activities:</td>
<td>6h</td>
<td>4.00%</td>
</tr>
<tr>
<td></td>
<td>Self study:</td>
<td>96h</td>
<td>64.00%</td>
</tr>
</tbody>
</table>
**Content**

### Unconstrained Optimization

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

**Description:**
- Line search. Acceptability of step sizes.
- General minimization algorithm.
- Newton's method. Factorizations to ensure convergence.
- Weighted least squares.
- Introduction to AMPL. The Neos solver site.

### Constrained Optimization and Support Vector Machines.

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

**Description:**
- Introduction to Support Vector Machines (SVM)
- KKT Optimality conditions of constrained optimization. Optimality conditions of SVM.
- Duality in Optimization. The dual of the SVM.

### Integer Programming

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

**Description:**
- Modelling problems with binary variables.
- The branch and bound algorithm for integer programming
- Gomory's cutting planes algorithm.
- Minimal spanning tree problem and algorithms of Kruskal and Prim.
## Planning of activities

### Unconstrained Optimization

**Description:**
- Optimality conditions.
- Convexity.
- Descent directions.
- Line search.
- Acceptability of step sizes.
- General minimization algorithm.
- Gradient method.
- Rate of convergence.
- Newton's method.
- Factorizations to ensure convergence.
- Weighted least squares.
- Introduction to AMPL. The Neos solver site.

**Specific objectives:**
1, 2, 3, 4

**Hours:** 52h 18m  
Theory classes: 17h 18m  
Practical classes: 0h  
Laboratory classes: 0h  
Guided activities: 0h  
Self study: 35h

### Constrained Optimization and Support Vector Machines

**Description:**
- Introduction to Support Vector Machines (SVM)
- KKT Optimality conditions of constrained optimization.
- Optimality conditions of SVM.
- Duality in Optimization. The dual of the SVM.

**Specific objectives:**
1, 2, 3, 4

**Hours:** 52h 24m  
Theory classes: 17h 24m  
Practical classes: 0h  
Laboratory classes: 0h  
Guided activities: 0h  
Self study: 35h

### Integer Programming

**Description:**
- Modelling problems with binary variables.
- The branch and bound algorithm for integer programming.
- Gomory’s cutting planes algorithm.
- Minimal spanning tree problem and algorithms of Kruskal and Prim

**Specific objectives:**
1, 2, 3, 4

**Hours:** 52h 18m  
Theory classes: 17h 18m  
Practical classes: 0h  
Laboratory classes: 0h  
Guided activities: 0h  
Self study: 35h
270659 - OTDM - Optimization Techniques for Data Mining

Qualification system

- Theory (40%). There will be a short midterm exam based on practical questions for the first two parts of the course, and some individual numerical exercises for the third (and last) part.

- Practical assignments (60%). There will be 3 lab assignments, one for each part of the course, all of them with the same weight on the final mark.

Additional exercises provided during the lectures may be taken into consideration to decide or to boost the final mark.

Bibliography

Basic:


Others resources:

Hyperlink


http://www-eio.upc.es/teaching/ple/pfc_ing.html