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

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

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>Hours large group: 24h</th>
<th>16.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hours medium group:</td>
<td>12h</td>
</tr>
<tr>
<td></td>
<td>Hours small group:</td>
<td>12h</td>
</tr>
<tr>
<td></td>
<td>Guided activities:</td>
<td>6h</td>
</tr>
<tr>
<td></td>
<td>Self study:</td>
<td>96h</td>
</tr>
</tbody>
</table>

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

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>Hours large group: 24h</th>
<th>16.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hours medium group:</td>
<td>12h</td>
</tr>
<tr>
<td></td>
<td>Hours small group:</td>
<td>12h</td>
</tr>
<tr>
<td></td>
<td>Guided activities:</td>
<td>6h</td>
</tr>
<tr>
<td></td>
<td>Self study:</td>
<td>96h</td>
</tr>
</tbody>
</table>
## Content

### Unconstrained Optimization

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

**Description:**
- 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.

### 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:**
- 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

<table>
<thead>
<tr>
<th>Hours: 52h 18m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 17h 18m</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: 35h</td>
</tr>
</tbody>
</table>

### 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

<table>
<thead>
<tr>
<th>Hours: 52h 24m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 17h 24m</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: 35h</td>
</tr>
</tbody>
</table>

### 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

<table>
<thead>
<tr>
<th>Hours: 52h 18m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 17h 18m</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: 35h</td>
</tr>
</tbody>
</table>
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

https://neos-server.org/neos/

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