ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

93000938 - Optimization techniques in big data analysis

DEGREE PROGRAMME

09AT - Master Universitario En Teoria De La Señal Y Comunicaciones

ACADEMIC YEAR & SEMESTER

2018/19 - Semester 1

Index

Learning guide

1. Description

1.1. Subject details

Name of the subject	93000938 - Optimization techniques in big data analysis
No of credits	3 ECTS
Type	Optional
Academic year ot the programme	First year
Semester of tuition	Semester 1
Tuition period	September-January
Tuition languages	English
Degree programme	09AT - Master universitario en teoria de la señal y comunicaciones
Centre	09 - Escuela Tecnica Superior de Ingenieros de Telecomunicacion
Academic year	2018-19

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
Santiago Zazo Bello (Subject coordinator)	C-326	santiago.zazo@upm.es	Sin horario. Arrange the meeting by email
Jesus Grajal De La Fuente	C-420	jesus.grajal@upm.es	Sin horario. Arrange the meeting by email

* The tutoring schedule is indicative and subject to possible changes. Please check tutoring times with the faculty member in charge.

3. Prior knowledge recommended to take the subject

3.1. Recommended (passed) subjects

- Fundamentos de optimización

3.2. Other recommended learning outcomes

- Statistical Signal Processing

4. Skills and learning outcomes *

4.1. Skills to be learned

CB06 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación

CB07 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CB08 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CB09 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo

CE01 - Analizar y aplicar técnicas para el diseño y desarrollo avanzado de equipos y sistemas, basándose en la teoría de la señal y las comunicaciones, en un entorno internacional

CE03 - Valorar y contrastar la utilización de las diferentes técnicas disponibles para la resolución de problemas reales dentro del área de teoría de la señal y comunicaciones.

CT01 - Capacidad para comprender los contenidos de clases magistrales, conferencias y seminarios en lengua inglesa

CT03 - Capacidad para adoptar soluciones creativas que satisfagan adecuadamente las diferentes necesidades planteadas

CT04 - Capacidad para trabajar de forma efectiva como individuo, organizando y planificando su propio trabajo, de forma independiente o como miembro de un equipo

CT05 - Capacidad para gestionar la información, identificando las fuentes necesarias, los principales tipos de documentos técnicos y científicos, de una manera adecuada y eficiente

4.2. Learning outcomes

RA14 - Capability to model real phenomena using probability theory.

RA2 - Capacidad para planificar, diseñar y realizar aplicaciones que integren técnicas de tratamiento de señal, análisis estadístico y aprendizaje automático sobre datos masivos.

RA17 - Capacidad para aplicar conocimientos de modelado estadístico, técnicas de optimización y modelos de series temporales en el análisis de datos y como base para el desarrollo de algoritmos de aprendizaje automático

RA15 - Capability to relate the foundations of statistical inference with standard machine learning schemes.

RA18 - Knowledge of tools for description, analysis and modeling of discrete-time random processes

RA25 - Handle with ease the bases of linear algebra and calculus necessary to formulate problems optimization.

RA7 - Capacidad para desarrollar y evaluar técnicas de aprendizaje automático y diseñar sistemas de aprendizaje para datos masivos

RA1 - Capacidad para desarrollar técnicas de tratamiento de señal específicas para datos masivos y diseñar aplicaciones sobre señales como: imágenes, señales de video, voz, audio y las procedentes de sensores de diversanaturaleza

RA32 - Capability for planning, design and implement applications, incorporating signal processing, statistical

analysis and machine learning

RA13 - Capability to construct parameter estimators, hypothesis tests and linear regression models.

RA19 - Knowledge of tools to design optimal filtering and signal processing structures

RA20 - Capability to choose the appropriate modeling and filtering tools in order to extract useful information from a time series

RA26 - Ability of oral and written communication

RA12 - Capability to construct probabilistic models from experimental data using inference tools.

RA6 - Saber resolver problemas de optimización con o sin restricciones mediante métodos analíticos y numéricos

RA5 - Saber resolver problemas de optimización básicos como los de programación lineal o cuadrática

RA4 - Formular problemas relacionados con la ingeniería como problemas de optimización en forma estándar

* The Learning Guides should reflect the Skills and Learning Outcomes in the same way as indicated in the Degree Verification Memory. For this reason, they have not been translated into English and appear in Spanish.

5. Brief description of the subject and syllabus

5.1. Brief description of the subject

This subject uses most of the topics already provided in Optimization Fundamentals now focused on the specific problem arising with massive data. Although we will provide the theoretical foundations on the evolved techniques we will also emphasize different case studies in big data applications. We will distinguish three main blocks

1. Fundamentals of Machine Learning as optimization problems and dimensionality reduction techniques

2. Local and efficient first and second order methods

3. Parallelization and in-network optimization techniques with massive data

5.2. Syllabus

1. Introduction

2. Machine Learning Contextualization

3. Review of Fundamentals of Convex Optimization

4. Data Dimensionality Reduction

5. First Order Methods

6. Second Order Methods. Quasi-Newton implementations

7. Augmented Lagrangian methods

8. ADMM: a Distributed Implementation using Concensus Principles

9. In Network Optimization

10. Parallel Optimization for Convex / Non Convex Functions

6. Schedule

6.1. Subject schedule*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Other face-to-face activities	Assessment activities
1				
2				
3				
4				
5				
6				
7				
8	**Fundamentals of Large Scale Optimization. Introduction and machine Learning contextualization** Duration: 02:00 Lecture	**Personal work related to chapter 2** Duration: 01:00 Laboratory assignments		**Personal work programming an optimization problem related to chapter 2** Problem-solving test Continuous assessment Duration: 01:00
9	**Review of fundamentals of convex optimization** Duration: 02:00 Lecture	**Personal work related to chapter 3** Duration: 01:00 Laboratory assignments		**Personal work programming an optimization problem related to chapter 3** Problem-solving test Continuous assessment Duration: 01:00
10	**Data Dimensionality Reduction. Large Scale Optimization. Algorithms for local processing in parallel settings. First Order Methods** Duration: 02:00 Lecture	**Personal work related to chapter 5** Duration: 01:00 Laboratory assignments		**Personal work programming an optimization problem related to chapter 5** Problem-solving test Continuous assessment Duration: 01:00
11	**Large Scale Optimization. Algorithms for local processing in parallel settings. Second Order Methods** Duration: 01:00 Lecture	**Personal work related to chapter 6** Duration: 01:00 Laboratory assignments		**Personal work programming an optimization problem related to chapter 6** Problem-solving test Continuous assessment Duration: 01:00 **Mid term exam corresponding to chapters 1-5. The student will have to solve an optimization problem equivalent to the laboratory activities** Problem-solving test Continuous assessment Duration: 01:00
12	**Augmented Lagrangian Methods Parallelized / Distributed Large Scale Convex Optimization. ADMM: a Distributed implementation using Consensus principles.** Duration: 01:00 Lecture	**Personal work related to chapters 7,8** Duration: 02:00 Laboratory assignments		**Personal work programming an optimization problem related to chapters 7,8** Problem-solving test Continuous assessment Duration: 01:00

13	**In Network optimization** Duration: 02:00 Lecture	**Personal work related to chapter 9** Duration: 01:00 Laboratory assignments		**Personal work programming an optimization problem related to chapter 9** Problem-solving test Continuous assessment Duration: 01:00
14	**Parallel optimization for convex / non convex functions** Duration: 02:00 Lecture	**Personal work related to chapter 10** Duration: 01:00 Laboratory assignments		**Personal work programming an optimization problem related to chapter 10** Problem-solving test Continuous assessment Duration: 01:00
15				
16				
17				**Final exam. The student will have to solve an optimization problem equivalent to the laboratory activities covering all the chapters. The student has to provide all the reports corrsponding to the exercises of the whole course** Problem-solving test Final examination Duration: 02:00 **Second exam. The student will have to solve an optimization problem equivalent to the laboratory activities** Problem-solving test Continuous assessment Duration: 02:00

The independent study hours are training activities during which students should spend time on individual study or individual assignments.

Depending on the programme study plan, total values will be calculated according to the ECTS credit unit as 26/27 hours of student face-to-face contact and independent study time.

* The subject schedule is based on a previous theorical planning of the subject plan and might go to through experience some unexpected changes along throughout the academic year.

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Continuous assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
8	Personal work programming an optimization problem related to chapter 2	Problem-solving test	Face-to-face	01:00	5%	0 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CT04 CE01 CE03 CT05 CB10
9	Personal work programming an optimization problem related to chapter 3	Problem-solving test	Face-to-face	01:00	5%	0 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CT04 CE01 CE03 CT05 CB10
10	Personal work programming an optimization problem related to chapter 5	Problem-solving test	Face-to-face	01:00	5%	0 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CT04 CE01 CE03 CT05 CB10

11	Personal work programming an optimization problem related to chapter 6	Problem-solving test	Face-to-face	01:00	5%	0 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CT04 CE01 CE03 CT05 CB10
11	Mid term exam corresponding to chapters 1-5. The student will have to solve an optimization problem equivalent to the laboratory activities	Problem-solving test	Face-to-face	01:00	30%	0 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CT04 CE01 CE03 CT05 CB10
12	Personal work programming an optimization problem related to chapters 7,8	Problem-solving test	Face-to-face	01:00	10%	0 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CT04 CE01 CE03 CT05 CB10
13	Personal work programming an optimization problem related to chapter 9	Problem-solving test	Face-to-face	01:00	5%	0 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CT04 CE01 CE03 CT05 CB10
14	Personal work programming an optimization problem related to chapter 10	Problem-solving test	Face-to-face	01:00	5%	0 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CT04 CE01 CE03

							CT05 CB10
17	Second exam. The student will have to solve an optimization problem equivalent to the laboratory activities	Problem-solving test	Face-to-face	02:00	30%	0 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CT04 CE01 CE03 CT05 CB10

7.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Final exam. The student will have to solve an optimization problem equivalent to the laboratory activities covering all the chapters. The student has to provide all the reports corrsponding to the exercises of the whole course	Problem-solving test	Face-to-face	02:00	100%	0 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CT04 CE01 CE03 CT05 CB10

7.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
Final exam. The student will have to solve an optimization problem equivalent to the laboratory activities It will cover all the chapters. The student has to provide all the reports corrsponding to the exercises of the whole	Written test	Face-to-face	03:00	100%	0 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CT04 CE01 CE03

course						CT05 CB10

7.2. Assessment criteria

Students will be qualified through continuous evaluation by default. According to the Normativa de Evaluación del Aprendizaje de la Universidad Politécnica de Madrid, students willing to renounce to continuous evaluation must complete the Moodle task entitled "Renounce to continuous evaluation" before the first intermediate exam (deadline will be announced in Moodle).
Evaluation will assess if students have acquired all the competences of the subject. Thus, evaluation through final assessment will be carried out considering all the evaluation techniques used in continuous evaluation (EX, ET, TG, etc.), and will be celebrated in the exam period approved by Junta de Escuela for the current academic semester and year. Evaluation activities that assess learning outcomes that cannot be evaluated through a single exam can be carried out along the semester.
Extraordinary examination will be carried out exclusively by the final assessment method.

The evaluation procedure for the continuous assessment will be as follows:

- After each chapter, the student will have to program an optimization problem in Matlab similar to the case study described by the instructor. After completing the program and running some simulations, a short report has to be written and sent to the instructor for evaluation. Each report will be scored with 5% of the final mark
- One mid term exam including the first 5 chapters counting 30% of the final mark has to be completed. The student will have to solve a certain number of theoretical / practical issues similar as those contents of the practices and lectures.
- One second term exam including the last chapters from 6 to 10 counting 30% of the final mark has to be completed. The student will have to solve a certain number of theoretical / practical issues similar as those contents of the practices and lectures.

The evaluation procedure for the final and re-sit examination will be as follows:

- A short report including the code and derivations for every exercise in the course has to be written and sent to the instructor for evaluation. Each report will be scored with 5% of the final mark.
- A final exam counting 60% of the final mark has to be completed. The student will have to solve a certain number of theoretical / practical issues similar as those contents of the practices and lectures.

8. Teaching resources

8.1. Teaching resources for the subject

Name	Type	Notes
Optimization techniques for Big Data analysis	Bibliography	Notes describing all the contents of the course
Matlab code of case studies	Others	Matlab code is provided for solving all the case studies proposed in the course
Course slides	Bibliography	Slides to be presented by the instructor to support the explanations
Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Foundations and Trends in Machine Learning, 3(1):1-122, 2011.	Bibliography	Main reference of chapters 7 and 8
Sayed, A. H. Adaptation, learning, and optimization over networks. Foundations and Trends in Machine Learning, vol.7, no.4-5, pp. 311-801, 2014.	Bibliography	Main reference chapter 9
Sayed, A. H. and Tu, Sheng-Yuan and Chen, J. and Zhao, X. and Towfic, Z. J, Diffusion strategies for adaptation and learning over networks. IEEE Signal Processing Magazine, vol. 30, no 3, pp.155-171, May 2013.	Bibliography	Important reference chapter 9

Sayed, A. H. , Adaptive Networks. Proceedings of the IEEE, vol. 102, no.4, pp.460-497, April 2014.	Bibliography	Important reference chapter 9
G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, Decomposition by Partial Linearization: Parallel Optimization of Multi-Agent Systems,? IEEE Transactions on Signal Processing, vol. 62, no. 3, pp. 641?656, Feb. 2014.	Bibliography	Important reference chapter 10
F. Facchinei, G. Scutari, and S. Sagratella, Parallel Selective Algorithms for Nonconvex Big Data Optimization, IEEE Transactions on Signal Processing, vol. 63, no. 7, pp. 1874-1889, Apr. 2015.	Bibliography	Important reference chapter 10
M. Hong, M. Razaviyayn, Z. Luo, and J. Pang, A Unified Algorithmic Framework for Block-Structured Optimization Involving Big Data: With applications in machine learning and signal processing. IEEE Signal Processing Magazine,	Bibliography	Important reference chapter 10