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

2023/24 - 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	2023-24

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
Juan Parras Moral	C-303	j.parras@upm.es	Sin horario. Arrange the meeting by email
Santiago Zazo Bello	C-326	santiago.zazo@upm.es	Sin horario. Arrange the meeting by email

Juan Ignacio Godino Llorente (Subject coordinator)		ignacio.godino@upm.es	- -

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

2.2. Research assistants

Name and surname	Email	Faculty member in charge
Arias Londoño, Julian David	julian.arias@upm.es	Godino Llorente, Juan Ignacio

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

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

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.

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

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

* 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 Large Scale Optimization

2. Large Scale Optimization. Algorithms for local processing in distributed settings

3. Bayesian optimization

5.2. Syllabus

1. Introduction

2. Machine Learning Contextualization

3. Review of Fundamentals of Convex Optimization

4. First Order Methods applied to Neural Networks

5. Second Order Methods

6. Coordinate Descent Methods

7. Augmented Lagrangian Methods

8. Symbolic differentiation and computational graphs

9. Bayesian optimization

6. Schedule

6.1. Subject schedule*

Week	Classroom activities	Laboratory activities	Distant / On-line	Assessment activities
1				
2				
3				
4				
5				
6				
7				
8	**1. Introduction 2. Machine Learning contextualization** Duration: 03:00 Lecture	**Personal work related to chapter 2** Duration: 01:00 Laboratory assignments		
9	**3. Review of fundamentals of convex optimization** Duration: 03:00 Lecture	**Personal work related to chapter 3** Duration: 01:00 Laboratory assignments		
10	**4. First Order Methods applied to Neural Networks** Duration: 03:00 Lecture	**Personal work related to chapter 4** Duration: 01:00 Laboratory assignments		
11	**5. Second Order Methods 6. Coordinate Descent Methods** Duration: 03:00 Lecture	**Personal work related to chapters 5 and 6** Duration: 01:00 Laboratory assignments		
12	**6. Augmented Lagrangian Methods** Duration: 02:00 Lecture			**Mid term exam corresponding to chapters 1-4. The student will have to solve problems equivalent to the laboratory activities covering chapters 1-4.** Problem-solving test Continuous assessment Presential Duration: 02:00
13	**8. Symbolic differentiation and computational graphs** Duration: 03:00 Lecture	**Personal work related to chapter 8** Duration: 01:00 Laboratory assignments		
14	**9. Bayesian optimization** Duration: 03:00 Lecture	**Personal work related to chapter 9** Duration: 01:00 Laboratory assignments		
15				

16				
17				**First exam. The student will have to solve problems equivalent to the laboratory activities covering chapters 1-4.** Problem-solving test Final examination Presential Duration: 02:00 **Second exam. The student will have to solve an optimization problem equivalent to the laboratory activities covering chapters 5-9.** Problem-solving test Continuous assessment and final examination Presential Duration: 02:00 **Deliver the final report with solved exercises** Individual work Continuous assessment and final examination Not Presential Duration: 00:00

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 schedule is based on an a priori planning of the subject; it might be modified during the academic year, especially considering the COVID19 evolution.

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
12	Mid term exam corresponding to chapters 1-4. The student will have to solve problems equivalent to the laboratory activities covering chapters 1-4.	Problem-solving test	Face-to-face	02:00	30%	3.5 / 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 covering chapters 5-9.	Problem-solving test	Face-to-face	02:00	30%	3.5 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CT04 CE01 CE03 CT05 CB10
17	Deliver the final report with solved exercises	Individual work	No Presential	00:00	40%	3 / 10	CB08 CT01 CB07 CT03 CB06 CT04 CE01 CE03 CT05 CB10

7.1.2. Global examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	First exam. The student will have to solve problems equivalent to the laboratory activities covering chapters 1-4.	Problem-solving test	Face-to-face	02:00	30%	3.5 / 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 covering chapters 5-9.	Problem-solving test	Face-to-face	02:00	30%	3.5 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CT04 CE01 CE03 CT05 CB10
17	Deliver the final report with solved exercises	Individual work	No Presential	00:00	40%	3 / 10	CB08 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 corresponding to the exercises of the whole course	Problem-solving test	Face-to-face	03:00	60%	3.5 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CT04 CE01 CE03 CT05 CB10
Deliver the final report with solved exercises	Individual work	Face-to-face	00:00	40%	3 / 10	CT01 CB07 CB08 CT03 CB06 CT04 CE01 CE03 CT05 CB10

7.2. Assessment criteria

Students will be qualified through progressive evaluation by default, and if they do not show up in the midterm exam, they will automatically be qualified using the final evaluation. Evaluation will assess if students have acquired all the competencies of the subject.

Evaluation through final assessment will be carried out considering all the evaluation techniques used in progressive evaluation (EX, ET, TG, etc.), and will take place during the exams 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 during the semester.
The extraordinary examination will be carried out exclusively by the final assessment method.

The evaluation procedure for the **progressive assessment** will be as follows:

- 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 to those contents of the practices and lectures. A minimum mark (3.5) is required.
- 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 to those contents of the practices and lectures. A minimum mark (3.5) is required.
- Both exams will include solving a short practical problem with the computer. The mark will be between 0 and 1 and will be considered an indicator to check if the student is able to solve problems individually.
- Report including all requested exercises. The mark between 0-4 will be weighted multiplicatively by the indicator described in the previous paragraph and the product will count 40% of the final mark.

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

- 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 to those contents of the practices and lectures.
- The final exam will include solving a short practical problem with the computer. The mark will be between 0 and 1 and will be considered an indicator to check if the student is able to solve problems individually.
- Report including all requested exercises. The mark between 0-4 will be weighted multiplicatively by the indicator described in the previous paragraph and the product will count 40% of the final mark

Extraordinary evaluation will be carried out exclusively by the final assessment method, just in the same way as the global evaluation in the ordinary call.

Academic fraud:

Any assessment or report may require a complementary oral assessment by the professor in order to validate that the task has been done by the student without help. According to the current assessment norms at UPM, if academic fraud is detected on any assessment, the student(s) will receive a grade of zero in the final grade of the examination to which the assessment belonged (ordinary or extraordinary).

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
Algorithms for case studies	Others	Source 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	
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	
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	
Sayed, A. H. , Adaptive Networks. Proceedings of the IEEE, vol. 102, no.4, pp.460-497, April 2014.	Bibliography	

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	
Neural Networks. A comprehensive Foundation. Simon haykin.IEEE Press 1994	Bibliography	
Optimization for Machine Learning, MIT Press, Edited by Suvrit Sra, Sebastian Nowozin and Stephen J. Wright, 2011	Bibliography	Main reading for the course
Numerical Optimization, J. Nocedal, S. Wright, Spinger, 2006	Bibliography	Basic booktext for optimization principles

9. Other information

9.1. Other information about the subject

This subject shows the fundamental ideas of distributed optimization that could be used to model different ODS objectives ,as biological models (ODS 3), climate changing (ODS 13) or ecosystems (ODSs 14 y 15). It could also be applied to the efficient use of resources as water (ODS 6) or energy (ODS 7)).
In more general terms, we teach applied mathematics used exhaustively in engineering, in particular will affect telecommunications infraestructures (ODS 9).

This course also contributes to subobjetives 4.4: to increase the number of persons with professional competences and techniques to access to employment and entrepeunership and 4.7, to guarantee that all students acquire solid practical and theoretical knowledge required to promote sustainable development