



POLITÉCNICA

INTERNATIONAL  
CAMPUS OF  
EXCELLENCE

COORDINATION PROCESS OF  
LEARNING ACTIVITIES  
PR/CL/001



E.T.S. de Ingenieros  
Informáticos

# ANX-PR/CL/001-01

## LEARNING GUIDE

### SUBJECT

**103000866 - Massively Parallel Machine Learning**

### DEGREE PROGRAMME

10AZ - Master Universitario En Innovación Digital

### ACADEMIC YEAR & SEMESTER

2024/25 - Semester 1

## Index

---

### Learning guide

|   |   |
|---|---|
| 1. Description.....                                   | 1 |
| 2. Faculty.....                                       | 1 |
| 3. Skills and learning outcomes .....                 | 2 |
| 4. Brief description of the subject and syllabus..... | 3 |
| 5. Schedule.....                                      | 4 |
| 6. Activities and assessment criteria.....            | 6 |
| 7. Teaching resources.....                            | 8 |
| 8. Other information.....                             | 9 |

## 1. Description

---

### 1.1. Subject details

|                                       |  |
|---------------------------------------|--|
| <b>Name of the subject</b>            | 103000866 - Massively Parallel Machine Learning          |
| <b>No of credits</b>                  | 4.5 ECTS   |
| <b>Type</b>                           | Optional   |
| <b>Academic year of the programme</b> | Second year  |
| <b>Semester of tuition</b>            | Semester 3   |
| <b>Tuition period</b>                 | September-January  |
| <b>Tuition languages</b>              | English  |
| <b>Degree programme</b>               | 10AZ - Master Universitario en Innovación Digital        |
| <b>Centre</b>                         | 10 - Escuela Tecnica Superior De Ingenieros Informaticos |
| <b>Academic year</b>                  | 2024-25  |

## 2. Faculty

---

### 2.1. Faculty members with subject teaching role

| <b>Name and surname</b>                              | <b>Office/Room</b> | <b>Email</b>  | <b>Tutoring hours *</b>                 |
|--|--------------------|---------------|---|
| Bonifacio Alberto Mozo Velasco (Subject coordinator) | 4307               | a.mozo@upm.es | W - 16:00 - 17:30<br>Th - 16:00 - 17:30 |

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

## 3. Skills and learning outcomes \*

---

### 3.1. 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

CG03 - La capacidad de usar la lengua inglesa de manera competente, es decir, con capacitación para tareas complejas de trabajo y estudio.

CG07 - Capacidad de trabajar y comunicarse también en contextos internacionales.

### 3.2. Learning outcomes

RA47 - Design massively parallel versions of supervised machine learning algorithms

RA44 - Understand the basic mechanisms for designing parallel applications in Big Data regimes

RA46 - Apply parallelization strategies to machine learning algorithms

\* 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.

## 4. Brief description of the subject and syllabus

---

### 4.1. Brief description of the subject

This course introduces to the parallelization of Machine Learning algorithms (supervised and unsupervised) on Big Data distributed computing platforms.

In particular, MapReduce and RDD paradigms are explained jointly with the corresponding distributed platforms supporting them (Hadoop and Apache Spark).

Using as examples several well-known machine learning algorithms, the student will learn how to parallelize machine learning algorithms using the corresponding Hadoop and Apache Spark APIs.

### 4.2. Syllabus

1. Introduction to Big Data
2. Machine Learning (supervised and unsupervised techniques)
3. MapReduce and Hadoop
4. RDDs and Spark
5. Parallelization of machine learning algorithms

## 5. Schedule

### 5.1. Subject schedule\*

| Week | Type 1 activities   | Type 2 activities   | Distant / On-line | Assessment activities  |
|------|---|---|-------------------|--|
| 1    | <b>Lecture on Unit 1</b><br>Duration: 02:00<br>Lecture                        |   |                   |  |
| 2    | <b>Lecture on Unit 2</b><br>Duration: 02:00<br>Lecture                        |   |                   |  |
| 3    | <b>Practical classes</b><br>Duration: 03:00<br>Problem-solving class          |   |                   |  |
| 4    | <b>Practical classes</b><br>Duration: 03:00<br>Problem-solving class          |   |                   |  |
| 5    | <b>Practical classes</b><br>Duration: 03:00<br>Problem-solving class          |   |                   |  |
| 6    | <b>Lecture on Unit 3</b><br>Duration: 02:00<br>Lecture                        |   |                   |  |
| 7    | <b>Lecture on Unit 4</b><br>Duration: 02:00<br>Lecture                        |   |                   |  |
| 8    | <b>Lecture on Unit 4</b><br>Duration: 02:00<br>Lecture                        |   |                   |  |
| 9    |   |   |                   | <b>Individual Exam</b><br>Written test<br>Progressive assessment and Global Examination<br>Presential<br>Duration: 02:00 |
| 10   | <b>Explanation of Project P1</b><br>Duration: 02:00<br>Laboratory assignments | <b>Project Development</b><br>Duration: 03:00<br>Cooperative activities |                   |  |
| 11   | <b>Explanation of Project P1</b><br>Duration: 02:00<br>Laboratory assignments | <b>Project Development</b><br>Duration: 03:00<br>Cooperative activities |                   |  |
| 12   | <b>Explanation of Project P2</b><br>Duration: 02:00<br>Laboratory assignments | <b>Project Development</b><br>Duration: 03:00<br>Cooperative activities |                   |  |
| 13   | <b>Explanation of Project P2</b><br>Duration: 02:00<br>Laboratory assignments | <b>Project Development</b><br>Duration: 03:00<br>Cooperative activities |                   |  |

|    |  |  |  |   |
|----|--|--|--|---|
| 14 |  |  |  | <b>Student Project Presentations P1</b><br>Group presentation<br>Progressive assessment and Global Examination<br>Presential<br>Duration: 02:00 |
| 15 |  |  |  | <b>Student Project Presentations P2</b><br>Group presentation<br>Progressive assessment and Global Examination<br>Presential<br>Duration: 02:00 |
| 16 |  |  |  |   |
| 17 |  |  |  |   |

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.

## 6. Activities and assessment criteria

### 6.1. Assessment activities

#### 6.1.1. Assessment

| Week | Description                      | Modality           | Type         | Duration | Weight | Minimum grade | Evaluated skills             |
|------|----------------------------------|--------------------|--------------|----------|--------|---------------|------------------------------|
| 9    | Individual Exam                  | Written test       | Face-to-face | 02:00    | 40%    | 5 / 10        | CB06<br>CB07<br>CG03<br>CG07 |
| 14   | Student Project Presentations P1 | Group presentation | Face-to-face | 02:00    | 30%    | 5 / 10        | CB06<br>CG03<br>CG07         |
| 15   | Student Project Presentations P2 | Group presentation | Face-to-face | 02:00    | 30%    | 5 / 10        | CG07<br>CB06<br>CG03         |

#### 6.1.2. Global examination

| Week | Description                      | Modality           | Type         | Duration | Weight | Minimum grade | Evaluated skills             |
|------|----------------------------------|--------------------|--------------|----------|--------|---------------|------------------------------|
| 9    | Individual Exam                  | Written test       | Face-to-face | 02:00    | 40%    | 5 / 10        | CB06<br>CB07<br>CG03<br>CG07 |
| 14   | Student Project Presentations P1 | Group presentation | Face-to-face | 02:00    | 30%    | 5 / 10        | CB06<br>CG03<br>CG07         |
| 15   | Student Project Presentations P2 | Group presentation | Face-to-face | 02:00    | 30%    | 5 / 10        | CG07<br>CB06<br>CG03         |

#### 6.1.3. Referred (re-sit) examination

| Description | Modality | Type | Duration | Weight | Minimum grade | Evaluated skills |
|-------------|----------|------|----------|--------|---------------|------------------|
|-------------|----------|------|----------|--------|---------------|------------------|



|  |                      |              |       |     |        |                              |
|--|----------------------|--------------|-------|-----|--------|------------------------------|
| The student will take an individual written exam that covers all units of the course     | Written test         | Face-to-face | 02:00 | 40% | 5 / 10 | CG03<br>CG07<br>CB07<br>CB06 |
| The student will hand practical assignments P1 and P2 up to the day of the written exam. | Problem-solving test | Face-to-face | 00:00 | 60% | 5 / 10 | CB06<br>CG03<br>CG07         |

## 6.2. Assessment criteria

### Assessment Criteria

#### Continuous assessment

The following partial grades will be obtained (grades from 0 to 10):

**N1: Grade for the individual exam.**

**P1: Grade for the student project P1**

**P2: Grade for the student project P2**

The final grade is obtained as follows:  $N = 0.40 \cdot N1 + 0.30 \cdot P1 + 0.30 \cdot P2$

As a general criterion, to pass the course (in the first or second call) it is required:

- $N1 \geq 5$  and  $P1$  and  $P2 \geq 5$ .

#### Final exam

The final grade is obtained as follows:  $N = 0.40 \cdot N1 + 0.30 \cdot P1 + 0.30 \cdot P2$

**N** is the grade obtained in the final exam

**P1:** Grade for the student project P1

**P2:** Grade for the student project P2

As a general criterion, to pass the course in the final exam it is required:

**N1  $\geq$  5 and P1 and P2  $\geq$  5.**

## 7. Teaching resources

---

### 7.1. Teaching resources for the subject

| Name                           | Type         | Notes  |
|--------------------------------|--------------|--|
| Resilient Distributed Datasets | Bibliography | Zaharia, Matei, et al. "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing." Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation. USENIX Association, 2012.                |
| MapReduce                      | Bibliography | Dean, Jeffrey, and Sanjay Ghemawat. "MapReduce: simplified data processing on large clusters." Communications of the ACM 51.1 (2008): 107-113.   |
| UPM Moodle Web Site            | Web resource | The web site UPM-Moodle ( <a href="http://moodle.upm.es">http://moodle.upm.es</a> ) will be used to provide course material to students. Students will use this web site to submit for evaluation the results of their assignments and individual works. |

|   |              |   |
|---|--------------|---|
| Machine Learning. A probabilistic approach. | Bibliography | Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press. |
|---|--------------|---|

## 8. Other information

---

### 8.1. Other information about the subject

Sustainable Development Goals:

This subject is aligned with Goal 4 Quality Education that aims to "Guarantee inclusive, equitable and quality education and promote lifelong learning opportunities for all" that defined within the Sustainable Development Goals (SDGs) established by the United Nations Organization ( UN).