ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

93000941 - Predictive And Descriptive Learning

DEGREE PROGRAMME

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

ACADEMIC YEAR & SEMESTER

2024/25 - Semester 1

Index

Learning guide

1. Description

1.1. Subject details

Name of the subject	93000941 - Predictive And Descriptive Learning
No of credits	6 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	2024-25

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
Eduardo Lopez Gonzalo (Subject coordinator)	C-330	eduardo.lopez@upm.es	Sin horario. Appointment arranged by email
Luis Alfonso Hernandez Gomez	C-330	luisalfonso.hernandez@upm.es	Sin horario. Appointment arranged by email

Juan Ignacio Godino Llorente	C-312	ignacio.godino@upm.es	Sin horario. Appointment arranged 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

The subject - recommended (passed), are not defined.

3.2. Other recommended learning outcomes

- It is mandatory to follow this course simultaneously with the subject Machine Learning Lab

- Previous exposure to a programming language, such as MATLAB, R or Python

- Elementary course in Statistics

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

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

CE02 - Evaluar y sintetizar los resultados de un trabajo en equipo en proyectos relacionados con 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

RA34 - Capability to develop and evaluate machine-learning techniques and to design big data learning systems

RA72 - Capability to understand, design, develop and evaluate machine-learning, deep-learning and generative AI technologies for a wide range of application areas

* 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 course covers the principles and methodology for the design, evaluation and selection of a large variety of Machine Learning, Deep Learning and AI technologies for practical applications. The course introduces main principles in Machine Learning: supervised, unsupervised and reinforcement learning, though main emphasis is on predictive and descriptive learning as reinforcement learning is covered in a subsequent course. Methodological issues such as model assessment and selection, and overfitting are discussed.

The course starts introducing the most relevant traditional predictive or supervised techniques: as different types of regression, generalized linear models, k-nearest neighbor classifier, classification and regression trees, ensemble methods (Bagging, Random Forests and Boosting) and kernel methods and Support Vector Machines. Then the course addresses traditional descriptive or unsupervised techniques: principal components analysis and clustering methods (k-means and hierarchical clustering). In parallel with classical ML, Deep Learning models (Feed-Forward Networks, Convolutional Networks, Recurrent Neural Networks and Transformers) are presented together with new ways of learning (self-supervised, contrastive, zero-shot, in-context, etc.). The course also introduces the main Generative AI models: Autoregressive, VAE, VQ-VAE, GANs, Flow Models, Diffusion models, Consistency Models. Emergent architectures such as Mamba, Kolmogorov?Arnold Networks, etc. are also discussed.

The students will understand the fundamentals and important topics in statistical machine learning, deep learning and artificial intelligence. This outcome represents a fundamental ingredient in the training of a modern AI engineering providing a solid base for its use on a wide range of applications in science and industry.

Through several examples and use cases, worked out in the companion course Machine Learning Lab, students will learn how important is to accurately assess the performance of a model. They will also acquire solid criteria on what could be best model for a given data and task. By the end of the course, students should be able to:

- Understand the fundamentals of the most used models and techniques for predictive and descriptive learning.
- Design a proper methodology for accurately assessing and gaining knowledge from the use of each one of the particular machine learning techniques.
- Know the strengths and weaknesses of the various approaches in order to choose the best models for a given data and application scenario.

5.2. Syllabus

1. Introduction to Machine Learning
    1.1. What is statistical learning?
    1.2. Types of Machine Learning
    1.3. Assessing Model Accuracy
2. Linear Regression
    2.1. Simple and Multiple Linear Regression
    2.2. Linear Regression and Distributed Machine Learning Principles
    2.3. Interpreting Regression Coefficients
    2.4. Model Selection and Qualitative Predictors
    2.5. Interactions and Nonlinearity
    2.6. Comparison of Linear Regression with KNN
3. Classification
    3.1. Logistic Regression
    3.2. Bayes classifier and Linear Discriminant Analysis
    3.3. Classification error analysis
    3.4. Quadratic Discriminant Analysis
    3.5. K-Nearest Neighbors
    3.6. A Comparison of Classification Methods: Logistic Regression, LDA, QDA and KNN
4. Resampling methods
    4.1. Cross-validation
    4.2. Bootstrap
5. Linear Model Selection and Regularization
    5.1. Feature selection
    5.2. Optimal Model selection
    5.3. Regularization
    5.4. Dimension Reduction
    5.5. High-Dimensional Data

6. Moving Beyond Linearity

   6.1. Generalized Linear Models and Generalized Additive Models

7. Tree-Based Methods

   7.1. Decision trees

   7.2. Bagging

   7.3. Random Forests

   7.4. Boosting

8. Support Vector Machines

   8.1. Maximal Margin Classifier

   8.2. Support Vector Classifiers

   8.3. Kernels and Support Vector Machines

   8.4. Relationship to Logistic Regression

9. Descriptive Learning

   9.1. Supervised vs Unsupervised learning

   9.2. Principal Components Analysis

   9.3. Clustering Methods

   9.4. K-means

   9.5. Hierarchical Clustering

   9.6. Practical Issues in Clustering

10. Introduction to Deep Learning

   10.1. Feed-forward Networks, Convolutional Networks

   10.2. Recurrent Networks and SeqSeq

   10.3. Attention Mechanisms and Transformers

   10.4. Ablation studies, Interpretability and Explainability

11. Introduction to Deep Generative Models

   11.1. Autoregressive Models: Large Language Models and Large Multimodal Models

   11.2. VAE, VQ-VAE

   11.3. GANs, Flow-based Models

   11.4. Diffusion Models and Consistency Models

12. Emergent models and new architectures

6. Schedule

6.1. Subject schedule*

Week	Type 1 activities	Type 2 activities	Distant / On-line	Assessment activities
1		**Activities Chapter 1** Duration: 02:00 Lecture **Activities Chapter 2** Duration: 02:00 Lecture		
2		**Activities Chapter 3** Duration: 04:00 Lecture		
3		**Activities Chapter 4** Duration: 02:00 Lecture **Activities Chapter 5** Duration: 01:00 Lecture **Activities Chapter 6** Duration: 01:00 Lecture		
4		**Activities Chapter 7** Duration: 01:00 Lecture **Activities Chapter 8** Duration: 03:00 Lecture		
5		**Activities Chapter 9** Duration: 04:00 Lecture		
6		**Activities Chapter 10 (10.1 , 10.2)** Duration: 04:00 Lecture		
7		**Activities Chapter 10 (10.3, 10.4)** Duration: 04:00 Lecture		
8		**Activities Use Case Review** Duration: 04:00 Cooperative activities		
9		**Activities Chapter 11 (11.1)** Duration: 04:00 Lecture		

10		**Activities Chapter 11 (11.2)** Duration: 04:00 Lecture		
11		**Activities Chapter 11 (11.3)** Duration: 04:00 Lecture		
12		**Activities Chapter 11 (11.4)** Duration: 03:00 Lecture **Activities Chapter 12** Duration: 01:00 Problem-solving class		
13		**Activities Use Case Review** Duration: 04:00 Cooperative activities		
14		**Activities Use Case Review** Duration: 04:00 Cooperative activities		
15				
16				
17				**Final project deep learning evaluation** Group work Progressive assessment Not Presential Duration: 00:00 **Evaluation: Machine Learning use case** Individual work Progressive assessment and Global Examination Not Presential Duration: 00:00 **Final project deep learning evaluation** Group work Global 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
17	Final project deep learning evaluation	Group work	No Presential	00:00	50%	3.5 / 10	CB09 CT01 CB07 CT03 CB06 CE02 CT04 CE01 CE03 CT05 CB10
17	Evaluation: Machine Learning use case	Individual work	No Presential	00:00	50%	3.5 / 10	CB09 CT01 CB07 CT03 CB06 CE02 CT04 CE01 CE03 CT05 CB10

7.1.2. Global examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Evaluation: Machine Learning use case	Individual work	No Presential	00:00	50%	3.5 / 10	CB09 CT01 CB07 CT03 CB06 CE02 CT04 CE01 CE03 CT05 CB10

17	Final project deep learning evaluation	Group work	No Presential	00:00	50%	3.5 / 10	CB09 CT01 CB07 CT03 CB06 CE02 CT04 CE01 CE03 CT05 CB10

7.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
Evaluation: Machine Learning use case	Individual work	Face-to-face	00:00	50%	3.5 / 10	CB09 CT01 CB07 CT03 CB06 CE02 CT04 CE01 CE03 CT05 CB10
Final project evaluation	Group work	Face-to-face	00:00	50%	3.5 / 10	CB09 CT01 CB07 CT03 CB06 CE02 CT04 CE01 CE03 CT05 CB10

7.2. Assessment criteria

Evaluation will assess if students have acquired all the competences of the subject. Thus, evaluation through extraordinary assessment will be carried out considering all the evaluation techniques used in ordinary evaluation (EX, ET, TG, etc.).

Progressive evaluation will be the preferred assessment method as it will be suited to the optimum learning process along the course. Progressive evaluation will consist of:

- A Machine Learning Report describing the activities that demonstrate skills in developing Machine Learning models. The evaluation of this Report will represent 50% of final grade. For progressive evaluation, this report must be due by the 9th week. Several course assignments, which will be announced in Moodle, will planned to review the students' progress through draft versions of their reports so we can give them feedback. We cloud also require students to prepare specific presentations to review their work.
- A Deep Learning Report must be prepared by the end of the course to demonstrate skills in developing Deep Learning models. The evaluation of this Deep Learning Report will represent 50% of final grade. Through several course assignments, announced in Moodle, we will review the students' progress while working in this Report. We cloud require students to attend to specific presentations to review their work. Deep Learning activities must be developed in working teams, but each team member must individually clearly describe her/his specific activities in the Report.

One group project subject could be in the form of a challenge in quantitative finance, which may include:

(1) Quantification of patterns in leading market indicators

(2) Quantification of patterns in market value.

(3) Quantification of seasonality in the main market indicators

(4) Quantification of intra- and inter-day patterns, volatility, etc.

Students in groups of two may request the challenge. The challenge statement will be published at the beginning of the course, including a calendar that will be in accordance

Students may request the challenge as a group or individually to the teacher responsible for the group. In the

latter case, the teachers will be in charge of forming the groups. In both scenarios the groups will be formed by

two students, and the students must meet the requirements to be able to develop the group work. The

challenge statement will be published at the beginning of the course, including a calendar that will be in accordance

with the rest of the course. The development of the challenge will be divided into four phases:

(1) Research: study of the challenge statement and research on possible solutions. The students will have to inform

themselves and formulate questions that will allow them to understand the dimension of the challenge and to

approach a possible solution.

(2) Development of the challenge: students will develop in teams small activities leading to identify the most

appropriate solution to the problem, all of them proposed by the teacher in view of the previous stages.

(3) Verification and validation: the results obtained and the chosen solution will be contrasted in real

environments.

(4) Elaboration of the report and/or exhibition: the results will be shared through a working report and/or an exhibition, which may be done through a video.

The monitoring of the phases of the activity will be developed in tutorial sessions with the teachers designated for this purpose. The evaluation will be carried out in a coordinated way between the teachers and the participants in the teams. The teachers will carry out a continuous evaluation of the performance and the achievement of the objectives set during the development of the challenge for each student. Likewise, after completing the challenge, students will perform a self-evaluation and a cross evaluation. The weight of the exercise in the grade will be the same as that assigned to the group work.

Final evaluation will consist of:

- A Machine Learning Report describing the activities that demonstrate skills in developing Machine Learning models. The evaluation of this Report will represent 50% of final grade and it must be due by the final exam date, although students can submit draft versions before that date can they can receive feedback on their work.
- A Deep Learning Report must be prepared to demonstrate skills in developing Deep Learning models. The evaluation of this Deep Learning Report will represent 50% of final grade and it must be due by the by the final exam date.

Deep Learning activities must be developed in working teams, but each team member must individually clearly

describe her/his specific activities in the Report.

For both Reports, students can submit draft versions before the final submission date so they can receive feedback

on their work.

Evaluation through extraordinary assessment will consist of:

- A Machine Learning Report describing the activities that demonstrate skills in developing Machine Learning

models (50% of final grade)

- A Deep Learning Report describing the activities that demonstrate skills in developing Deep Learning

models (50% of final grade)

8. Teaching resources

8.1. Teaching resources for the subject

Name	Type	Notes
An Introduction to Statistical Learning	Bibliography	James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. An introduction to statistical learning. second edition 2021 https://hastie.su.domains/ISLR2/ISLRv2_website.pdf

Machine learning: a probabilistic perspective	Bibliography	Kevin P. Machine learning: a probabilistic perspective. MIT press, 2012
The Elements of Statistical Learning Data Mining, Inference, and Prediction,	Bibliography	Hastie, Trevor, Tibshirani, Robert and Friedman, Jerome. The Elements of Statistical Learning Data Mining, Inference, and Prediction, Second Edition. Springer Series in Statistics, 2009
Deep learning	Bibliography	Goodfellow, I., Bengio, Y., Courville, A., &Bengio, Y. (2016). Deep learning. Cambridge: MIT press.
Neural Networks and Deep Learning	Web resource	http://neuralnetworksanddeeplearning.com/index.html
Scaling up machine learning: Parallel and distributed approaches.	Bibliography	Bekkerman, Ron, Mikhail Bilenko, and John Langford, eds. Scaling up machine learning: Parallel and distributed approaches. Cambridge University Press, 2011
Pattern recognition and machine learning (information science and statistics).	Bibliography	Christopher M. Bishop. Pattern Recognition and Machine Learning (Information Science and Statistics), 2006.
Hands-On Machine Learning with Scikit-Learn, Keras, and tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems	Bibliography	Géron, Aurélien. O'Reilly Media 2nd edition

9. Other information

9.1. Other information about the subject

For on-line learning activities we will use UPM Moodle platform and tools. Moodle, GutHub and Youtube will be the environments to share specific course materials.
The increasing relevance of technological developments based on Machine Learning makes this course an educational activity directed to contribute to Goal 4.4 in Sustainable Development Goals (SDGs) 2030 United Nations Agenda, empowering our students with relevant skills, including technical and vocational skills, for employment, decent jobs and entrepreneurship.

Through approaching practical scenarios in our Lab, students will develop relevant skills and in-depth knowledge

on the impact of different Machine Learning techniques on different fields as health, environmental monitoring,

smart energy management, or finance. This will help them to become more aware of how technology can

contribute to several SDGs goals: end poverty (Goal 1), promote well-being (Goal 2), and promote sustainable

management of water, energy, economic growth and industrialization (Goals 5, 6, 7, and 8) as well as to reduce

inequality among countries (Goal 10).