



FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Introducción al Análisis Estadístico con Lenguajes de Programación para Machine Learning
Código	DMA-MBD-516
Título	Máster Universitario en Big Data por la Universidad Pontificia Comillas
Impartido en	Máster Universitario en Big Data [Primer Curso] Máster Universitario en Ingeniería de Telecomunicación + Máster Universitario en Big Data [Primer Curso] Máster Universitario en Ingeniería de Telecomunicación + Máster Universitario en Big Data [Segundo Curso]
Nivel	Posgrado Oficial Master
Cuatrimestre	Semestral
Créditos	3,0 ECTS
Carácter	Obligatoria
Departamento / Área	Departamento de Organización Industrial
Responsable	José Portela González

Datos del profesorado	
Profesor	
Nombre	Pablo Carlos del Saz-Orozco Huang
Departamento / Área	Departamento de Métodos Cuantitativos
Correo electrónico	pcdelsazorozco@icai.comillas.edu
Profesor	
Nombre	Alejandro Polo Molina
Departamento / Área	Instituto de Investigación Tecnológica (IIT)
Correo electrónico	apolo@comillas.edu

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura
Aportación al perfil profesional de la titulación
<p>La asignatura Introducción al Análisis Estadístico con Lenguajes de Programación para Machine Learning proporciona al alumno una base esencial para el estudio avanzado de machine learning, al combinar conocimientos estadísticos con habilidades prácticas en programación. El curso equipa al estudiante con una comprensión sólida de métodos estadísticos y técnicas de modelado, junto con la capacidad para implementar estos métodos utilizando lenguajes de programación como Python. Al dominar conceptos avanzados de optimización, estimación de parámetros y modelado estadístico, el alumno está preparado para enfrentarse a problemas complejos y desarrollar soluciones efectivas en machine learning. Esta base robusta facilita el aprendizaje de asignaturas más avanzadas, asegurando que el estudiante tenga las competencias necesarias para aplicar técnicas sofisticadas y enfrentarse a desafíos en el campo de la ciencia de datos y la inteligencia artificial.</p>



Prerrequisitos

Se requiere conocimientos básicos de Cálculo y Álgebra (entender y manipular ecuaciones, manipular exponentes y logaritmos usando sus reglas básicas, comprensión completa de funciones y funciones inversas, entender límites, derivadas e integrales, conocer reglas para productos y sumas, etc.). Se recomienda encarecidamente, aunque no es obligatorio, tener conocimientos básicos de Estadística (estadísticas descriptivas, modelos de distribuciones de probabilidad discretas y continuas, muestreo y fundamentos de inferencia estadística).

También se requieren conocimientos básicos de lenguajes de programación, idealmente en Python.

Competencias - Objetivos

Competencias

Conocimientos o contenidos

CO1	Entender los fundamentos de la analítica de datos y su aplicación en diversas áreas de la inteligencia artificial, destacando la integración en soluciones complejas y multidisciplinarias para el análisis avanzado de datos masivos atendiendo a la diversidad de problemas específicos de cada área.
CO2	Comprender las técnicas de procesados de datos, las arquitecturas y herramientas más habituales y apropiadas para condiciones y requisitos de casos específicos.

Competencias

CP4	Implementar las técnicas de procesamiento de datos y usar las herramientas más habituales y apropiadas a las condiciones y requisitos de casos específicos.
------------	---

Habilidades o destrezas

HA1	Comunicar de manera oral y escrita con rigor técnico, claridad expositiva y coherencia argumentativa a todo tipo de interlocutores, técnicos y no técnicos.
HA3	Desarrollar las habilidades interpersonales que requieren los entornos profesionales actuales (empatía, tolerancia, respeto, capacidad para aunar intereses contrapuestos).
HA4	Gestionar, organizar y planificar adecuadamente el trabajo y el tiempo, cumpliendo objetivos y estándares de calidad.
HA5	Mantener una formación y aprendizaje continuo y adaptación a los cambios tecnológicos y científicos.

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

CONTENIDOS

1. Álgebra Lineal Avanzada y Optimización para Machine Learning:

- Descomposición en valores singulares (SVD) y Factorización de matrices
- Métodos de optimización avanzados como Newton-Raphson o métodos estocásticos



2.- Teoría de la Información:

- Entropía y entropía cruzada
- Información mutua
- Teorema del código fuente y del canal
- Divergencia de Kullback-Leibler (KL)

3.- Inferencia en modelos lineales y modelos lineales generalizados:

- Algoritmo Expectation-Maximization (EM) para estimación de parámetros
- Inferencia para modelos lineales generalizados (GLM)

4.- Estadística Bayesiana:

- Distribuciones a priori y a posteriori
- Cadenas de Markov Monte Carlo (MCMC)
- Muestreo de Gibbs
- Procesos gaussianos

5.- Procesos Estocásticos en Tiempo Discreto

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

Este curso tiene un enfoque práctico de “aplicar y entender de inmediato” en Estadística. Priorizamos la comprensión conceptual y utilizamos las herramientas computacionales en Python para hacer que los conceptos cobren vida en el aula y en la práctica de los estudiantes.

Metodología Presencial: Actividades

In-class activities

- **Clases:** El profesor introducirá los conceptos fundamentales de cada unidad, junto con algunas recomendaciones prácticas, y revisará ejemplos resueltos para apoyar la explicación. Se fomentará la participación activa mediante la formulación de preguntas abiertas para estimular la discusión y la propuesta de cuestionarios en línea y ejercicios cortos de aplicación que se resolverán en clase, ya sea en papel o utilizando un paquete de software
- **Sesiones de laboratorio:** Bajo la supervisión del instructor, los estudiantes, divididos en pequeños grupos, aplicarán los conceptos y técnicas abordados en las clases para recorrer las etapas del flujo de trabajo de análisis de datos.
- **Se organizarán tutorías para grupos o estudiantes individuales según se solicite.**



Metodología No presencial: Actividades

Out-of-class activities

- Estudio personal del material del curso y resolución de los ejercicios propuestos.
- Preparación de las sesiones de laboratorio para aprovechar al máximo el tiempo en clase.
- Análisis de los resultados del laboratorio y redacción de informes.
- Desarrollo de un proyecto final en grupos pequeños.

RESUMEN HORAS DE TRABAJO DEL ALUMNO

HORAS PRESENCIALES

HORAS NO PRESENCIALES

CRÉDITOS ECTS: 3,0 (0 horas)

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

El uso de IA para crear trabajos completos o partes relevantes, sin citar la fuente o la herramienta o sin estar permitido expresamente en la descripción del trabajo, será considerado plagio y regulado conforme al Reglamento General de la Universidad.

Calificaciones

Grading

Convocatoria ordinaria

- **Theory** will account for 50%, of which:
 - Midterm: 20%
 - Final exam: 30%
- **Lab** will account for the remaining 50%, of which:
 - Lab practices: 20%
 - Final project: 30%

Para aprobar el curso, la nota media ponderada debe ser igual o superior a 5 sobre 10 puntos, la nota del examen final debe ser igual o superior a 4 sobre 10 puntos, y la nota del laboratorio (la media ponderada de las tareas y el proyecto final) debe ser al menos 5 sobre 10 puntos. De lo contrario, la nota final será la más baja de las tres calificaciones.

Extraordinaria

Lab marks will be preserved as long as the weighted average of the assignments and the final project results in a passing grade. Otherwise, a new project will have to be developed and handed in. In addition, all students will take a final exam. The resulting

grade will be computed as follows:

- **Theory** will account for 50%, of which the best of the following two options will be taken:
 - Option 1:
 - Midterm: 20%
 - Final exam: 30%
 - Option 2:
 - Final exam: 50%
- **Lab** will account for the remaining 50%, of which:
 - If the student passed the lab during regular assessment
 - Lab assignments: 10%
 - Final project: 40%
 - Otherwise
 - Final project: 60%

Al igual que en el periodo de evaluación regular, para aprobar el curso, la nota media ponderada debe ser igual o superior a 5 sobre 10 puntos, la nota del examen final debe ser igual o superior a 4 sobre 10 puntos, y la nota del laboratorio debe ser al menos 5 sobre 10 puntos. De lo contrario, la nota final será la más baja de las tres calificaciones.

Reglas

La asistencia a clase es obligatoria de acuerdo con el Artículo 93 del Reglamento General de la Universidad Pontificia Comillas y el Artículo 6 de las Normas Académicas de la Escuela de Ingeniería ICAI. No cumplir con este requisito puede tener las siguientes consecuencias:

Los estudiantes que falten a más del 15% de las clases podrán perder el derecho a presentar el examen final durante el periodo de evaluación regular. En cuanto al laboratorio, la ausencia a más del 15% de las sesiones puede resultar en la pérdida del derecho a presentar el examen final del periodo de evaluación regular y del examen de recuperación. Las sesiones perdidas deben ser recuperadas para obtener el crédito correspondiente. Los estudiantes que cometan una irregularidad en cualquier actividad evaluada recibirán una calificación de cero en la actividad y se iniciará un procedimiento disciplinario (cf. Artículo 168 del Reglamento General de la Universidad Pontificia Comillas).

PLAN DE TRABAJO Y CRONOGRAMA

Actividades		Fecha de realización	Fecha de entrega
IN-CLASS ACTIVITIES			
LECTURE &			
Session/semana	h/s	PROBLEM	Contentts LAB ASSESMENT
SOLVING			
1	2	Introduction (1h)	Introduction to Machine learning - Technical Lab 1 - Intro python - GIT - Software
		Conceptos de entropía y entropía cruzada: Divergencia KL.	



2	2	Information Theory	Información de Fisher Definiciones, propiedades y ejemplos. Información mutua. Teorema del código fuente	Lab 2: Information Theory
3	2	Optimization	Métodos de optimización avanzados: Newton-Raphson y métodos estocásticos (SGD, Adam)	Lab 3: Optimization
4	2	Inference LM I	Logistic Regression, parameter estimation, Introducción al algoritmo Expectation-Maximization (EM) and MLE	Lab 4: Inference I
5	2	Inference LM II	Linear Regression: Parameter estimation MLE estimation and regularization	Lab 5: Inference II
6	2	Inference GLM	GLM: Fundamentals: estructura, función de enlace y ejemplos Inferencia en GLM: Estimación de parámetros, verosimilitud, bondad del ajuste	Lab 6: Inference GLM
7	2	Mid-term exam I		Mid-term exam I
8	2	Stochastic	Introduction Linear process - Random walk - ARMA	Lab 6.1: Stochastic
9	2	Stochastic	Markov Process	Lab 6.2: Markov Process
10	2	Algebra	EIG - SVD - Factorizacion	Lab 7: Factorization



11	2	Bayes I	Priori - Post MCMC	Lab 8.1	
12	2	Bayes III	Gibbs sampling	Lab 8.2	
13	2	Bayes IV	Gaussian process Introducción a procesos gaussianos: Definición, propiedades y aplicaciones en regresión no paramétrica	Lab 8.3	
14	2	Overview			Trabajo Final

BIBLIOGRAFÍA Y RECURSOS

Bibliografía Básica

Basic bibliography

- Slides and Jupyter notebooks prepared by the lecturers (available in Moodlerooms and the GitHub repository of this course).
- James G., Witten D., Hastie T., Tibshirani R., Taylor J. (2023) An Introduction to Statistical Learning with Applications in Python. <https://link.springer.com/book/10.1007/978-3-031-38747-0>
- Murphy, K. P. (2022). *Probabilistic machine learning: an introduction*. MIT press. Available at <https://probml.github.io/pml-book/book1.html>

Bibliografía Complementaria

Complementary bibliography

- Murphy, K. P. (2023). *Probabilistic machine learning: Advanced topics*. MIT press. Available at <https://probml.github.io/pml-book/book2.html>
- Kurt W. (2019) [Bayesian Statistics the Fun Way](#). No Starch Press. ISBN-13: