

Curso Académico: (2024 / 2025)

Fecha de revisión: 23-04-2024

Departamento asignado a la asignatura: Departamento de Teoría de la Señal y Comunicaciones

Coordinador/a: ARENAS GARCIA, JERONIMO

Tipo: Obligatoria Créditos ECTS : 6.0

Curso : 1 Cuatrimestre : 1

REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)

Ninguna (en relación a otras asignaturas del Máster), ya que es una asignatura del primer cuatrimestre. Sin embargo, es altamente aconsejable disponer de conocimientos básicos sobre estadística.

OBJETIVOS

Al finalizar el curso el alumno comprenderá la naturaleza de los problemas de regresión, clasificación y, en general, de análisis de datos, y conocerá distintas aproximaciones para su resolución. Tomará conciencia de la importancia que tiene en la comprensión de estos problemas el dominio de tres elementos básicos de la teoría de la probabilidad: la verosimilitud, la diferencia entre incertidumbre a priori y a posteriori y el teorema de Bayes.

Desde un punto de vista procedimental, el alumno abordará la resolución de estos problemas mediante diferentes técnicas de aprendizaje a partir de datos: técnicas de tipo no paramétrico, basadas en la minimización de riesgo empírico, o en el tratamiento bayesiano del problema.

De forma específica, los objetivos que se persiguen en la asignatura, enumerados como competencias, son los siguientes:

- capacitar al estudiante en los principios teóricos en que se basan distintas herramientas analíticas y algorítmicas básicas y avanzadas para el Tratamiento de Datos
- familiarizar al estudiante con la aplicación de dichas herramientas a casos reales y la extracción de conclusiones
- adquirir conocimiento y destreza en el manejo de herramientas clásicas de estimación y clasificación
- adquirir conocimiento y destreza en el manejo de herramientas de aprendizaje máquina: procesos gaussianos, máquinas de vectores soporte, métodos no paramétricos, etc.
- conocer otras aplicaciones del análisis de datos, en particular en el ámbito del procesamiento de lenguaje natural.

DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

Tema 0: Introducción general al tratamiento de datos

Tema 1: Preprocesamiento de datos

- 1.1. Normalización de datos
- 1.2. Reducción de dimensionalidad
- 1.3. Agrupamiento de datos

Tema 2: Regresión.

- 2.1. El problema de regresión
- 2.2. Regresión no paramétrica: k-NN
- 2.3. Regresión de mínimos cuadrados lineal y polinómica
- 2.4. Regresión bayesiana
- 2.5. Otros algoritmos de regresión

Tema 3: Clasificación.

- 3.1. Problema de clasificación.

- 3.2. Métodos no paramétricos: k-NN
- 3.3. Regresión logística
- 3.4. Otros algoritmos de clasificación

Tema 4: Deep learning

- 4.1. Introducción a pytorch
- 4.2. Redes neuronales de múltiples capas.
- 4.3. Redes convolucionales

Tema 5: Procesamiento de Lenguaje Natural

- 5.1. Preprocesamiento de textos.
- 5.2. Word embeddings
- 5.3. Transformers
- 5.4. Modelos de tópicos neuronales

ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

CLASES TEÓRICO - PRÁCTICAS

Serán sesiones magistrales en las que se presentarán los conceptos básicos de la asignatura, ilustrándolos con numerosos ejemplos. Se dedicarán también numerosas sesiones a la resolución de ejercicios y problemas de carácter similar a los que se plantearán en los exámenes.

PRÁCTICAS DE LABORATORIO

Sesiones en aula informática de aplicación de los conceptos presentados en la asignatura. El alumno resolverá con ayuda del ordenador problemas de tratamiento de datos reales, evaluando las prestaciones de los sistemas implementados.

SISTEMA DE EVALUACIÓN

Peso porcentual del Examen Final:	50
Peso porcentual del resto de la evaluación:	50

La evaluación será continua. El total de la calificación de la asignatura se distribuye del siguiente modo:

- * Evaluación continua: 55%
 - Prueba de Laboratorio en aula: 25%
 - Proyecto de análisis de datos: 30%

- * Examen Final: 45 %

La convocatoria extraordinaria consistirá en tres pruebas: proyecto final, examen de teoría y examen de laboratorio. El alumno puede conservar cualesquiera de las calificaciones de la convocatoria ordinaria, pero la presentación a cualquiera de estas pruebas implica renunciar automáticamente a la nota correspondiente de la convocatoria ordinaria.

BIBLIOGRAFÍA BÁSICA

- Denis Rothman Transformers for Natural Language Processing, Packt>, 2022 (2nd ed)
- R. O. Duda, P. E. Hart, D. G. Stork Pattern Classification (2nd ed.), Wiley Interscience, 2001

BIBLIOGRAFÍA COMPLEMENTARIA

- A.C. Müller, S. Guido Introduction to Machine Learning with Python: A Guide for Data Scientists, O'Reilly Media, Inc, 2016
- C. M. Bishop Pattern Recognition and Machine Learning, Springer, 2006

- Trevor Hastie, Robert Tibshirani, Jerome H. Friedman The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer Series in Statistics, 2009

RECURSOS ELECTRÓNICOS BÁSICOS

- Aston Zhang and Zachary C. Lipton and Mu Li and Alexander J. Smola . Dive into Deep Learning: <https://d2l.ai/>
- J. Cid-Sueiro, J. Arenas-García, V. Gómez-Verdejo . Machine Learning 4 All: <https://github.com/ML4DS/ML4all>