

# Guía docente

## 270962 - AMA - Análisis Multivariante Avanzado

Última modificación: 23/11/2023

**Unidad responsable:** Facultad de Informática de Barcelona  
**Unidad que imparte:** 715 - EIO - Departamento de Estadística e Investigación Operativa.

**Titulación:** MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS (Plan 2021). (Asignatura optativa).

**Curso:** 2023      **Créditos ECTS:** 6.0      **Idiomas:** Inglés

### PROFESORADO

---

**Profesorado responsable:** PEDRO FRANCISCO DELICADO USEROS

**Otros:** Primer quadrimestre:  
PEDRO FRANCISCO DELICADO USEROS - 10

### COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

---

#### Específicas:

- CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa
- CE13. Identificar las principales amenazas en el ámbito de la ética y la privacidad de datos en un proyecto de ciencia de datos (tanto en el aspecto de gestión como de análisis de datos) y desarrollar e implantar medidas adecuadas para mitigar dichas amenazas.
- CE3. Aplicar métodos de integración de datos para dar solución a problemas de ciencia de datos en entornos heterogéneos
- CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos
- CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.
- CE8. Extraer información de datos estructurados y no estructurados, teniendo en cuenta la naturaleza multivariante de los mismos.

#### Genéricas:

- CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

#### Transversales:

- CT4. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN: Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de la especialidad y valorar de forma crítica los resultados de esta gestión.
- CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

#### Básicas:

- CB10. 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.
- CB6. 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.
- CB7. 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



## METODOLOGÍAS DOCENTES

---

Hay dos sesiones semanales de 2 horas.

Las tres primeras horas se dedican a la exposición de los temas teóricos por parte del profesor.

La última hora se dedica a implementar estos contenidos: Cada alumno tiene su portátil en clase y realiza las tareas propuestas por el profesor.

Cada semana finaliza con una tarea para los alumnos que debe ser entregada en 7 días. El software utilizado será principalmente R.

## OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

---

1. Conocer la estructura de los principales problemas de aprendizaje no supervisado.
2. Aprenda diferentes métodos para la reducción de la dimensionalidad cuando no se cumplen los supuestos estándar en el análisis multivariado clásico
3. Aprenda a combinar técnicas de reducción de dimensionalidad con algoritmos de predicción
4. Al finalizar el curso el alumno será capaz de proponer, estimar, interpretar y validar versiones no paramétricas de modelos de regresión lineal y modelos lineales generalizados.
5. Al finalizar el curso el alumno sabrá adecuadamente cómo elegir los parámetros de suavizado que en los modelos de regresión no paramétricos controlan el equilibrio entre un buen ajuste a la muestra observada y una buena generalización.
6. Al finalizar el curso, el alumno será consciente de la necesidad de dotar de interpretabilidad a los algoritmos de aprendizaje automático, conocerá las técnicas de interpretabilidad más habituales, sabrá cómo clasificarlas y qué relaciones existen entre ellas, y sabrá usarlos en R y/o Python.

## HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

---

Tipo	Horas	Porcentaje
Horas aprendizaje autónomo	96,0	64.00
Horas grupo grande	54,0	36.00

**Dedicación total:** 150 h

## CONTENIDOS

---

Aprendizaje no supervisado mediante análisis multivariado avanzado

Modelos de regresión no paramétricos

Aprendizaje automático interpretable



## ACTIVIDADES

### Aprendizaje no supervisado mediante el análisis multivariante avanzado

#### Objetivos específicos:

1, 2, 3

#### Competencias relacionadas:

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CE8. Extraer información de datos estructurados y no estructurados, teniendo en cuenta la naturaleza multivariante de los mismos.

CE3. Aplicar métodos de integración de datos para dar solución a problemas de ciencia de datos en entornos heterogéneos

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa

CT4. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN: Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB10. 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.

CB6. 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.

#### Dedicación: 52h 18m

Grupo grande/Teoría: 18h

Aprendizaje autónomo: 34h 18m



## Modelos de regresión no paramétrica

### Objetivos específicos:

4, 5

### Competencias relacionadas:

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CE8. Extraer información de datos estructurados y no estructurados, teniendo en cuenta la naturaleza multivariante de los mismos.

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa

CT4. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN: Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB10. 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.

**Dedicación:** 54h 18m

Grupo grande/Teoría: 20h

Aprendizaje autónomo: 34h 18m

## Interpretabilidad en aprendizaje automático

### Objetivos específicos:

6

### Competencias relacionadas:

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE13. Identificar las principales amenazas en el ámbito de la ética y la privacidad de datos en un proyecto de ciencia de datos (tanto en el aspecto de gestión como de análisis de datos) y desarrollar e implantar medidas adecuadas para mitigar dichas amenazas.

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CE8. Extraer información de datos estructurados y no estructurados, teniendo en cuenta la naturaleza multivariante de los mismos.

CT4. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN: Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB7. 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

**Dedicación:** 43h 18m

Grupo grande/Teoría: 16h

Aprendizaje autónomo: 27h 18m



## SISTEMA DE CALIFICACIÓN

---

A lo largo del curso se asignarán tareas como deberes. Las notas de los deberes valdrán el 40% de la nota del curso.

Habrà un examen al final del cuatrimestre y se evaluarà la asimilaci3n de los conceptos b3sicos de toda la asignatura. El examen final tendrà una primera parte te3rica breve (libros cerrados) y una segunda parte pr3ctica m3s larga (libros abiertos, a realizar por los alumnos con su propio ordenador port3til, con una estructura similar a los deberes).

Nota del curso =  $0.4 * \text{Nota Hwk} + 0.6 * \text{Nota del examen}$

## BIBLIOGRAFÍA

---

### B3sica:

- Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. The Elements of statistical learning : data mining, inference, and prediction. 2nd ed. New York [etc.]: Springer, cop. 2009. ISBN 9780387952840.
- Wasserman, Larry. All of nonparametric statistics. Pittsburgh: Springer, cop. 2010. ISBN 9781441920447.
- Wood, Simon N. Generalized additive models : an introduction with R. 2nd ed. Boca Raton: CRC Press/Taylor & Francis Group, [2017]. ISBN 9781498728331.
- Biecek, P.; Burzykowski, T. Explanatory model analysis: explore, explain and examine predictive models. Oxford: Oxford University Press, 2018. ISBN 9780367135591.
- Fan, Jianqing; Li, Runze; Zhang, Cun-hui; Zou, Hui. Statistical foundations of data science. Boca Raton: Oxon : CRC Press, 2020. ISBN 9781466510845.