



## Guía docente

# 200642 - ODS - Optimización en Ciencia de Datos

Última modificación: 29/05/2025

**Unidad responsable:** Facultad de Matemáticas y Estadística

**Unidad que imparte:** 715 - EIO - Departamento de Estadística e Investigación Operativa.

**Titulación:** MÁSTER UNIVERSITARIO EN ESTADÍSTICA E INVESTIGACIÓN OPERATIVA (Plan 2013). (Asignatura optativa).

**Curso:** 2025

**Créditos ECTS:** 5.0

**Idiomas:** Castellano, Inglés

## PROFESORADO

**Profesorado responsable:** JORDI CASTRO PÉREZ

**Otros:** Primer cuatrimestre:

DANIEL BAENA MIRABETE - A  
JORDI CASTRO PÉREZ - A

## CAPACIDADES PREVIAS

Conceptos básicos de estadística y de optimización/investigación operativa.

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

### Específicas:

6. CE-2. Capacidad para dominar la terminología propia de algún ámbito en el que sea necesaria la aplicación de modelos y métodos estadísticos o de investigación operativa para resolver problemas reales.
7. CE-3. Capacidad para formular, analizar y validar modelos aplicables a problemas de índole práctica. Capacidad de seleccionar el método y/o técnica estadística o de investigación operativa más adecuado para aplicar dicho modelo a cada situación o problema concreto.
8. CE-5. Capacidad para formular y resolver problemas reales de toma de decisiones en los diferentes ámbitos de aplicación sabiendo elegir el método estadístico y el algoritmo de optimización más adecuado en cada ocasión.
9. CE-6. Capacidad para utilizar el software más adecuado para realizar los cálculos necesarios en la resolución de un problema.
10. CE-7. Capacidad para comprender artículos de estadística e investigación operativa de nivel avanzado. Conocer los procedimientos de investigación tanto para la producción de nuevos conocimientos como para su transmisión.
11. CE-8. Capacidad de discutir la validez, el alcance y la relevancia de estas soluciones y saber presentar y defender sus conclusiones.
12. CE-9. Capacidad para implementar algoritmos de estadística e investigación operativa.



**Transversales:**

1. EMPRENDIMIENTO E INNOVACIÓN: Conocer y entender la organización de una empresa y las ciencias que rigen su actividad; tener capacidad para entender las normas laborales y las relaciones entre la planificación, las estrategias industriales y comerciales, la calidad y el beneficio.
2. SOSTENIBILIDAD Y COMPROMISO SOCIAL: Conocer y comprender la complejidad de los fenómenos económicos y sociales típicos de la sociedad del bienestar; tener capacidad para relacionar el bienestar con la globalización y la sostenibilidad; lograr habilidades para utilizar de forma equilibrada y compatible la técnica, la tecnología, la economía y la sostenibilidad.
3. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.
4. 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 especialidad, y valorar de forma crítica los resultados de dicha gestión.
5. 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.

## METODOLOGÍAS DOCENTES

---

Teoría:

Se presentan y discuten los contenidos de la asignatura combinando explicaciones en la pizarra y transparencias.

Prácticas:

Sesiones de laboratorio en que se muestra el uso de software.

## OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

---

El objetivo del curso es introducir al alumno en algunas aplicaciones en "data science" que pueden ser formuladas o solucionadas por técnicas de optimización. La asignatura tiene tres partes, cada una representa aproximadamente 1/3 del total:

1. Solución de problemas estadísticos y de aprendizaje automático usando optimización entera y combinatoria: diseño de experimentos (cuadrados latinos ortogonales), clustering óptimo (k-medoid), clustering heurístico (k-means).
2. Métodos de optimización continua para la solución de problemas de aprendizaje automático: regresión, "support vector machines" y redes neuronales.
3. Introducción al campo del control de la revelación estadística o protección de datos estadísticos: métodos para microdatos y tablas de datos. Esta disciplina propone un conjunto de métodos para garantizar la confidencialidad de datos individuales en diseminar datos estadísticos, sean microdatos o datos agregados en forma tabular. Este problema es de gran importancia para Institutos Nacionales de Estadística, y, en general, cualquier entidad privada u organismo oficial que tenga que divulgar datos.

Capacidades a adquirir:

- \* Formular problemas en "data science" como problemas de optimización (clustering, support vector machines, redes neuronales ...)
- \* Saber solucionar los problemas de "data science" formulados usando software de optimización y de aprendizaje automático (scikit-learn, tensorflow).
- \* Saber qué es el campo del control de la revelación estadística o protección de datos estadísticos.
- \* Conocer software para protección de datos y ser capaz de proteger datos usando alguna técnica existente.
- \* Familiarizarse con la literatura de optimización en "data science".



## HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas grupo grande	30,0	24.00
Horas grupo pequeño	15,0	12.00
Horas aprendizaje autónomo	80,0	64.00

**Dedicación total:** 125 h

## CONTENIDOS

### Optimización entera y combinatoria en problemas estadísticos y de "data science".

**Descripción:**

Conceptos básicos de optimización entera y combinatoria. Modelización de problemas de optimización entera y combinatoria. Aplicaciones: diseño de experimentos (cuadrados latinos ortogonales), clustering óptimo (k-medoids), clustering heurístico (k-means).

**Dedicación:** 15h

Grupo grande/Teoría: 10h

Grupo mediano/Prácticas: 5h

### Optimización continua en problemas de data science.

**Descripción:**

Regresión ridge y LASSO. Support Vector machines (SVMs): formulación primal, condiciones de optimalidad KKT, dualidad en SVMs, formulación dual, métodos de optimización para SVMs, software para SVMs. Redes neuronales: estructura y modelización de RNs como problema de optimización, métodos de optimización para RNs, software para RNs (scikit-learn, tensorflow).

**Dedicación:** 15h

Grupo grande/Teoría: 5h

Grupo mediano/Prácticas: 10h

### Introducción a la protección de datos estadísticos.

**Descripción:**

Introducción. Definiciones. Tipos de datos y métodos. Métodos de protección para microdatos. Métodos de protección para datos tabulares. Software de protección de datos.

**Dedicación:** 15h

Grupo grande/Teoría: 10h

Grupo mediano/Prácticas: 5h

## SISTEMA DE CALIFICACIÓN

Un examen parcial de la primera parte de la asignatura (40% de la nota) y realización de trabajos prácticos (60% de la nota)



## BIBLIOGRAFÍA

---

### Básica:

- Cristianini, Nello; Shawe-Taylor, John. An introduction to support vector machines and other kernel-based learning methods. Cambridge: Cambridge University Press, 2000. ISBN 0521780195.
- Willenborg, Leon; Waal, Ton de. Elements of statistical disclosure control. New York: Springer, 2001. ISBN 0387951210.
- Arthanari, T.S. Mathematical programming in statistics. Wiley, 1993. ISBN 0471592129.
- Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron. Deep learning [en línea]. Cambridge, Massachusetts: The MIT Press, [2016] [Consulta: 05/07/2023]. Disponible a: <https://www.deeplearningbook.org/>. ISBN 9780262035613.
- Nocedal, Jorge; Wright, Stephen J. Numerical optimization [en línea]. New York: Springer, 1999 [Consulta: 05/07/2023]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-40065-5>. ISBN 9780387987934.
- Luenberger, David G; Ye, Yinyu. Linear and nonlinear programming [en línea]. 3rd ed. New York: Springer, cop. 2008 [Consulta: 05/07/2023]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-74503-9>. ISBN 0387745033.