

**DEPARTAMENTO DE MATEMÁTICAS**  
**OFRECIMIENTOS DE CURSOS ELECTIVOS 202520**

<p align="center"><b>Nivel del Curso*</b></p> <p>4: posgrado      _x_</p> <p>3: final de carrera   _x_</p> <p>2: mitad de carrera   __</p> <p>1: inicio de carrera   __</p>	<p><b>Nombre completo del curso en español:</b></p> <p>Introducción matemática a la ciencia de datos</p>
	<p><b>Nombre completo del curso en inglés:</b></p> <p>Mathematical introduction to data sciences</p>
	<p><b>Nombre abreviado en español (Máx. 30 caracteres contando espacios)</b></p> <p>Ciencia de datos matemática</p>
	<p><b>Profesor:</b></p> <p>Michael A. Hoegele</p>
<p><b>Descripción del curso en español:</b></p> <p>Mientras en gran parte de la historia de la ciencia el progreso científico tenía como limitante la escasez generalizada de datos confiables, se invirtió este paradigma desde los años noventa del siglo pasado. Fomentado por cada vez mejores sensores, el crecimiento exponencial de la potencia computacional y el aumento de la capacidad de almacenamiento de datos la extracción de la información contenida en el diluvio de datos a disposición se ha vuelto un distintivo de nuestra era de épocas anteriores. En particular, se emancipó la disciplina de la ciencia de datos de la estadística clásica propia con algoritmos y preguntas enfocados con dimensiones muy altas y la eficiencia de esos algoritmos.</p> <p>Este curso cubre las técnicas y métodos clásicos de la ciencia de datos desde una perspectiva matemática con énfasis a los mecanismos detrás de cada uno y por tanto sus posibles adaptaciones junto con sus limitaciones. Las áreas matemáticas que se usan son teoría de probabilidad, estadística matemática, álgebra lineal, análisis funcional y optimización. Sin embargo, aparte de las bases matemáticas de teoría de probabilidad los resultados que se usan se introducen a lo largo del curso. Habrá también implementaciones y el uso de librerías estándares por los proyectos entregados y presentados en clase al final del curso, en particular, para los participantes de código 4.</p>	
<p><b>Descripción del curso en inglés:</b></p>	

While for much of the history of science scientific progress was limited by the widespread scarcity of reliable data, this paradigm has been reversed since the 1990s. Fueled by ever better sensors, exponential growth in computing power and increasing data storage capacity, extracting information from the flood of available data has become a hallmark of our era from previous eras. In particular, the discipline of data science has been emancipated from classical statistics with algorithms and questions focused on very high dimensions and the efficiency of those algorithms.

This course covers classical data science techniques and methods from a mathematical perspective with emphasis on the mechanisms behind each and therefore their possible adaptations along with their limitations. The mathematical areas used are probability theory, mathematical statistics, linear algebra, functional analysis and optimization. However, apart from the mathematical foundations of probability theory, the results used are introduced throughout the course. There will also be implementations and the use of standard libraries for the projects handed in and presented in class at the end of the course, in particular for the code 4 participants.

**Prerrequisitos:**

Probabilidad de honores o introducción a la probabilidad equivalente con un contenido matemático suficiente

**Objetivos:**

- Dominar rigurosamente los mecanismos matemáticos detrás de los algoritmos en la ciencia de datos.
- Analizar y evaluar las hipótesis de los resultados y métodos matemáticos, su rol y sus limitaciones
- Desarrollar una intuición matemática sobre posibles modificaciones de los métodos a nuevos contextos de aplicaciones
- Discutir y presentar el contexto de aplicaciones estándares en el lenguaje matemático adecuado
- Implementar los algoritmos usando en un entorno idealizado (en particular, por los proyectos)

**Contenido:**

- Regresión lineal, polinomial, y logística

- k vecinos más próximos
- Clustering, clustering sobre grafos
- Subespacios mejor adaptados
- Análisis de componentes principales
- Reducción de dimensionalidad
- Filtrado colaborativo
- Perceptrón
- Máquinas de vectores de soportes
- Método Kernel
- Descenso del gradient
- Redes neuronales

#### **Forma de Evaluación:**

##### **Código 3**

3 Parciales: 45%, un 15% c.u.

9 Tareas: 30%, un 3,33% c.u.

Asistencia: 5%, el porcentaje la asistencia confirmada en clase multiplicada con 5

15 Quices: 15%, cada semana inicia con un quiz de 15min que vale un 1% c.u.

1 Proyecto final: 5%.

##### **Código 4**

Se hacen las evaluaciones del código 3 y que equivalen proporcionalmente a un 80%,

Sin embargo el proyecto final de código 3 se reemplaza por 2 proyectos que equivalen en total a un 20% de la nota final (5% y 15% resp.)

3 Parciales:  $0.8 \cdot 45\%$ , un  $0.8 \cdot 15\%$  c.u.

9 Tareas:  $0.8 \cdot 30\%$ , un  $0.8 \cdot 3,33\%$  c.u.

Asistencia:  $0.8 \cdot 5\%$ , el porcentaje la asistencia confirmada en clase

multiplicada con 5  
15 Quices:  $0.8 \cdot 15\%$ , cada semana inicia con un quiz de 15min que vale un 1% c.u.

**Bibliografía principal:**

**Wegner: Mathematical introduction to data sciences, Springer, 2024.**

**<https://link.springer.com/book/9783662694251>**

**Bibliografía secundaria:**

**Vershynin: High-Dimensional Probability, vol 47 of Cambridge University press, 2018**

**Blum, Hopcroft, Kannan: Foundations of Data Sciences, Cambridge University Press, 2020**

**Shao: Mathematical statistics, Springer 2003**

**Klenke: Probability theory, Springer 2008**