

# Introducción a ciencia de datos

Agosto-Diciembre 2023

**Profesor:** Emilien Joly (ext. 531, of. H104, emilien.joly@cimat.mx)

**Ayudante:** A definir (xxx@cimat.mx)

**Horarios de clase:** A definir, 11:00-12:20

**Horarios de prácticas:** a definir, 11:00-12:20

## Objetivos

1. Introducir los conceptos clásicos de aprendizaje maquina, clasificación a agrupación de datos.
2. Evaluar la pertinencia y la calidad de algoritmos de aprendizaje maquina en función del contexto.
3. Aplicar los algoritmos del curso a algunos ejemplos de datos reales.
4. Familiarizarse con los algoritmos modernos de aprendizaje maquina.

## Temario

1. Métodos de clasificación
  - 1.1. Clasificador óptimo de Bayes
  - 1.2. Clasificador Bayesiano ingenuo
  - 1.3. Análisis discriminante
  - 1.4. Algoritmo de  $k$ -vecinos más cercanos
  - 1.5. Modelos de regresión
  - 1.6. Árboles de decisión
  - 1.7. Modelos gráficos no-dirigidos I: Modelos gaussianos e independencia condicional
2. Métodos de búsqueda de estructura
  - 2.1. Descomposición en valores singulares (SVD)
  - 2.2. Reducción de dimensión
  - 2.3. Análisis de componentes principales (PCA)
  - 2.4. Factorización nonegativa (NMF) y sistemas de recomendación
  - 2.5. Elementos de álgebra lineal aleatorizada
  - 2.6. Proyecciones aleatorias y el Lema de Johnson-Lindenstrauss
3. Agrupamiento de datos
  - 3.1. Análisis de conglomerados
  - 3.2. Métodos espectrales para agrupamientos y caminatas aleatorias
  - 3.3. Algoritmo EM
  - 3.4. Modelos de mezclas gaussianas
4. Técnicas de aprendizaje máquina para el análisis de datos
  - 4.1. Métodos kernel
  - 4.2. Máquinas de soporte vectorial
  - 4.3. Redes neuronales