

Introducción al Data Mining

Curso Optativo – A

Docente

Prof. Dr. Ricardo Fraiman
CMAT, Facultad de Ciencias

Reunión inicial: Lunes 4 de Abril 9:00hs. en el Institut Pasteur Montevideo en la cual se fijará el horario del curso

Objetivos

Acercar al estudiante a los principales métodos estadísticos en análisis multivariado. En particular, para clasificación supervisada y no supervisada (cluster analysis), componentes principales, selección de variables y otros métodos de reducción de dimensiones.

Conocimientos Previos

Probabilidad & estadística, álgebra lineal y conocimientos de R.

Temario

1. **Componentes principales.**
2. **Regresión no paramétrica.** Métodos basados en vecinos más cercanos. Métodos basados en núcleos.
3. **Clasificación supervisada.** Método de vecinos más cercanos.
4. **Clasificación supervisada.** Métodos jerárquicos y árboles binarios. CART.
5. **Clasificación supervisada en el caso infinito dimensional.** Extensiones del Teorema de Stone. Métodos de vecinos más cercanos y núcleos.
6. **Clasificación no supervisada.** Análisis cluster. k-medias y k-medioides.
7. **Clasificación no supervisada.** k-medias recortadas.
8. **Clasificación no supervisada por árboles binarios.** CUBT.
9. **Selección de variables para clasificación supervisada y no supervisada.**
10. **Reducción de dimensión a la Johnson-Lindenstrauss.**
11. **Algunas aplicaciones.**