

Métodos multivariados

Instituto Tecnológico Autónomo de México

Mauricio Benjamín García Tec

Primavera 2017

Objetivo

Este curso contiene una serie de temas selectos de Estadística Computacional, Análisis Multivariado de Datos y de Métodos Geométricos y Topológicos para Datos. Aunque todos los ejemplos de clase se harán en **R**, los estudiantes pueden usar **Python**, **Julia** o el lenguaje de su preferencia para implementar los algoritmos de clase. Sobre todo en la parte de Análisis Topológico de Datos, es importante que los estudiantes contribuyan al desarrollo de una paquetería de **R** que subiremos al CRAN.

Temario

1. **Visualizando datos multivariados.** Vamos a repasar algunos conceptos probabilísticos como varianza, covarianza y correlación y haremos gráficas y manipulaciones básicas usando **dplyr**, **ggplot2** y **plotly**. Se espera que los alumnos ya tengan alguna familiaridad con varias de estas herramientas. A lo largo del curso vamos a presentar algunas herramientas nuevas usando **javascript**, **D3** y **shiny** que los estudiantes podrán incluir en sus proyectos.
2. **PCA y reducción lineal de dimensionalidad.** Veremos con cuidado el álgebra lineal detrás del Análisis de Componentes Principales y veremos sus usos para visualizar datos, crear clusters, construir índices, etc.
3. **Métodos Numéricos para Análisis de Datos.** Durante el curso vamos a requerir algunas herramientas de optimización cuadrática. Presentaremos los básicos de optimización y en particular de optimización cuadrática con restricciones.
4. **Métodos de Kernel.** Introduciremos el truco del Kernel que será uno de los ejes del curso. Daremos una revisión del álgebra lineal detrás de los métodos del Kernel. Como aplicaciones veremos la reducción no lineal de dimensionalidad usando Kernel PCA y las máquinas de soporte vectorial. Veremos varias aplicaciones a problemas de clasificación.
5. **Otros métodos de clusterización.** Veremos algunos de los algoritmos populares de clusterización de datos como **dbscan**, **k-nearest neighbours** y un algoritmo basado en cambios de densidad que construiremos nosotros.
6. **Análisis Topológico de Datos.** Combinaremos todo lo aprendido en los módulos anteriores en una paquetería de visualización de datos con un enfoque topológico. Habrá proyectos finales para crear la documentación y tutoriales de nuestra paquetería.

Referencias

Muchos de los temas tratados en el curso son nuevos, por lo que las notas de clase serán la fuente principal. Para los temas de Métodos Numéricos y de Kernel daremos referencias bibliográficas en el momento.