

Taller: Técnicas de Clustering de datos

Director del taller: Dr. Cristian Martínez

Duración: 3 horas.

Destinatarios: alumnos avanzados, egresados y docentes de carrera de Informática y Matemática.

Conocimientos previos: métodos numéricos, estadística y programación.

Fundamentación

La Minería de Datos es un campo de las Ciencias de la Computación referido al proceso de descubrimiento de patrones en grandes volúmenes de datos. Para ello, utiliza técnicas y herramientas provenientes de la inteligencia artificial, aprendizaje automático, estadística y bases de datos.

Dentro de los métodos más conocidos se encuentran los de Clustering, los cuales son procedimientos algorítmicos que agrupan una colección de objetos de datos según criterios, habitualmente de distancia y/o densidad.

El taller está centrado en introducir el proceso de minería de datos y aplicar diferentes técnicas de Clustering sobre los datos procesados, aplicando herramientas específicas y de uso en la academia como SPSS y R.

Objetivo general

- Adquirir conocimientos teóricos y prácticos análisis de datos y técnicas de Clustering.

Objetivos específicos

- Analizar datos a nivel multidimensional.
- Conocer diferentes técnicas de Clustering.
- Extraer conocimiento a partir de los datos bajo estudio.
- Manejar herramientas actuales de Data Mining.

Contenidos

- Análisis multivariado de datos
 - Exploración de datos
 - Resumen de datos
 - Gráficos multivariados
- Métodos de Clustering
 - Jerárquicos
 - Particionales
 - Basado en densidad
 - Introducción a Text Mining
- Práctica en laboratorio usando R y SPSS

Metodología de Enseñanza

Las clases serán teórico-prácticas y serán desarrolladas en laboratorio. Los contenidos teóricos serán desarrollados mediante diapositivas y pizarrón. Las prácticas se realizarán usando computadoras.

Infraestructura y equipamiento necesarios

Las clases se dictarán preferentemente en laboratorio de computadoras y se usarán R y SPSS. Para el dictado de clases, se hará uso de Data Display y pizarrón.

Bibliografía básica

- Chandra, J., A survey on clustering algorithms for data in spatial database management systems. *International Journal of Computer Applications*, 24(9), 2011.
- Han, J., Kamber, M., *Data Mining: Concepts and Techniques*, Morgan Kauffmann, 2011.
- Johnson, D., *Métodos Multivariados Aplicados al Análisis de Datos*, Ed. THOMSON 1999.
- Witten, I., Elbe, F., *Data Mining: Practical Machine Learning Tools and techniques*, Morgan Kauffmann, 2011.
- Zhao, Y., *R and Data Mining*, Academic Press, 2012.