

MINERÍA DE DATOS: APRENDIZAJE NO SUPERVISADO Y DETECCIÓN DE ANOMALÍAS

Isaac Triguero Velázquez

Jesús Alcalá Fernández (jalcala@decsai.ugr.es)

Daniel Sánchez Fernández

Juan Carlos Cubero Talavera



Horarios y Aulas

□ Clases:

- Martes 7 Nov. (18:00-20:30), Miércoles 8 Nov. (18:00-20:30), Jueves 9 Nov. (18:00-19:15)
- Lunes 27 Nov. (15:30-18:00)

□ Despacho:

- D20 del departamento de CCIA (ETSIIT)
- Despacho M1 del Módulo B del Edf. Mecenass (Ciencias)

□ Tutorías (con cita previa – Módulo B de Mecenass):

- Lunes: 09:30-11:00
- Martes: 09:00-11:00
- Viernes: 09:30-11:00, 12:00-13:00

Temario de Teoría

□ Tema 1: Clustering

- Introducción, tipos de datos y distancias. Métodos jerárquicos. Métodos basados en centroides Métodos de evaluación: Interna y externa. Extensiones de las técnicas conocidas

□ Tema 2: Problemas de asociación entre atributos

- **Introducción a las reglas de asociación.**
- **Métodos de generación clásicos.**
- **Conjuntos maximales y cerrados.**
- **Tratamiento de valores numéricos.**
- **Reglas Multinivel.**
- **Evaluación: Medidas de interés**

□ Tema 3: Problemas de detección de anomalías

- Introducción al problema. Métodos Supervisados Métodos Semi-Supervisados Métodos No Supervisados

Prácticas

- Las prácticas se realizarán con **RStudio**.
- Objetivos:
 - ▣ Conocer el paquete “arules” de R.
 - ▣ Aplicar algunos de los métodos clásicos de extracción de reglas de asociación a varios problemas: Apriori y Eclat
 - ▣ Extracción de itemsets maximales y cerrados.
 - ▣ Estudio de las reglas según sus medidas de calidad.
 - ▣ Visualización de reglas de asociación (paquete “arulesViz”)
 - ▣ Introducción a **RKEEL**, una herramienta que integra un módulo avanzando para la extracción de reglas de asociación.

Organización

- Sesión 1:
 - **Introducción a las reglas de asociación.**
 - **Métodos de generación clásicos.**
 - **Conjuntos maximales y cerrados.**
- Sesión 2:
 - **Generación de reglas.**
 - **Problemas abiertos.**
 - **Aplicaciones.**
 - **Implementación con arules y arulesViz - caso de estudio (I).**
- Sesión 3:
 - **Implementación con arules y arulesViz - caso de estudio (II).**
 - **Extracción de RAs a partir de la base de datos Zoo.**

Evaluación Ordinaria

- Examen escrito global de la asignatura el día 27 de Dic (15:30).
- Un trabajo práctico a desarrollar por el alumno de aquella parte de la asignatura (clustering, anomalías, asociación) que el estudiante elija.
- Para RAs, dicho trabajo consistirá en aplicar lo aprendido en el guión de prácticas a un dataset distinto al utilizado en clase, a elegir por el estudiante. El trabajo debe incluir un análisis descriptivo de los resultados obtenidos.
- Reparto de estudiantes equitativo entre temáticas. La selección se puede hacer el 24 de Noviembre de 2023 a partir de las 22:00

Ponderación:

- Evaluación de la parte de teoría: 60%
- Evaluación de la parte de prácticas: 40%
- Necesario sacar en cada parte como mínimo un 3.5 sobre 10 para hacer media de las dos partes.

Evaluación Extraordinaria

- La evaluación sigue el mismo esquema que el examen de la convocatoria ordinaria, salvo que, si el alumno ya se hubiese presentado al examen escrito o entregado la práctica, se le mantiene la nota de la parte que el alumno decida (siempre que no se presente a esa parte en esta convocatoria extraordinaria).
- Se aplican los mismos porcentajes que los indicados en la convocatoria ordinaria.

Bibliografía y Enlaces de Interés

□ Bibliografía:

- ▣ Pang-Ning Tan, Michael Steinbach , Vipin Kumar. Introduction to Data Mining. Addison Wesley (2006)
- ▣ J. Han, M. Kamber, and J. Pei. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers Inc., 3rd edition, 2011.
- ▣ H.I. Rhys. Machine Learning with R, the tidyverse, and mlr. Manning, 2020.

□ Enlaces de Interés:

- ▣ RStudio (<http://www.rstudio.org/download/>)
- ▣ KEEL (<http://www.keel.es>) y RKEEL (<https://cran.r-project.org/web/packages/RKEEL/index.html>)

