



Laboratorio 3

Reglas de Asociación

Curso: Análisis de Datos

Profesor: Dr. Max Chacón

Ayudante: Marcelo Álvarez

9 de noviembre de 2024

1. Objetivos

- Extraer conocimiento de los conjuntos de datos asignados utilizando el algoritmo Apriori para descubrir patrones de asociación y realizar el análisis respectivo.
- Comparar los resultados obtenidos con lo expuesto en la literatura, evaluando si las reglas de asociación identificadas sustentan el conocimiento adquirido previamente en el análisis de datos.
- Analizar las reglas generadas e identificar aquellas asociaciones más relevantes que permitan inferir conocimiento a partir de los patrones observados, evaluando las métricas de soporte, confianza, y otras medidas de calidad.

2. Metodología y Recursos

- Utilizar R para el desarrollo de este laboratorio, específicamente el paquete *arulesViz* para obtener, visualizar y analizar reglas de asociación.
- Aplicar técnicas de preprocesamiento de datos, considerando la eliminación de datos redundantes o irrelevantes que puedan afectar el análisis de reglas de asociación. En el caso de variables numéricas, evaluar si es necesario aplicar discretización para convertirlas en categóricas; y para variables categóricas, considerar el uso de *one-hot encoding* para transformarlas en variables binarias. Justificar todas las decisiones de preprocesamiento realizadas.
- Aplicar el algoritmo Apriori (función `apriori`) para extraer conocimiento de la base de datos asignada, descubriendo patrones de asociación entre los datos. En este algoritmo, se debe configurar valores mínimos de confianza y de soporte para obtener reglas interesantes.
- Seleccionar las medidas de calidad adecuadas (e.g., soporte, confianza, lift) para evaluar las reglas generadas. Justificar la elección de estas métricas en función de la naturaleza de los datos y los objetivos específicos del análisis.

- Evaluar la monotonía y las propiedades de las medidas seleccionadas, analizando cómo estos factores afectan la interpretación y selección de las reglas de asociación.
- Comparar los resultados obtenidos con los de laboratorios anteriores.
- Toda información relevante de expertos proveniente de fuentes externas, debe ser debidamente citada en **APA 7**. Para ello, se recomienda utilizar la plantilla de informe disponible en el Campus Virtual, la cual viene configurada para utilizar APA mediante los comandos `\parencite` y `\textcite`. Las referencias en formato BibTeX se deben colocar en el archivo “bibliografia.bib”.

3. Estructura del Informe

- **Introducción (5 puntos):** Visión clara y concisa del estudio, incluyendo el contexto, los objetivos y un resumen de la metodología.
- **Marco Teórico (10 puntos):** Definir los conceptos utilizados en el laboratorio (por ejemplo, medidas de calidad, soporte, confianza, lift, etc.).
- **Preprocesamiento (15 puntos):** Se deben definir criterios para eliminar registros o columnas que contengan datos irrelevantes o que no aporten información útil para el análisis de reglas de asociación. Si se decide discretizar variables numéricas o aplicar técnicas de transformación como *one-hot encoding*, justificar dichas decisiones en función de su impacto en la calidad y relevancia de las reglas generadas.
- **Obtención de Reglas (15 puntos):** Describir el proceso utilizado para aplicar el algoritmo de asociación, incluyendo la configuración de parámetros como soporte mínimo, confianza mínima y cualquier otro parámetro relevante. Justificar la elección de estos parámetros en función del análisis. una descripción de los criterios utilizados para evaluar la calidad de los clústers obtenidos.
- **Análisis de Resultados (30 puntos):** Analizar las reglas de asociación obtenidas e identificar aquellas que resulten más interesantes o significativas en el contexto del

problema.. Contrastar estas reglas con lo expuesto en la literatura y comparar los hallazgos con los resultados de los laboratorios anteriores.

- **Conclusiones (15 puntos):** Concluir respecto a los resultados obtenidos, el desarrollo del laboratorio y el método utilizado. Mencionar aspectos positivos y a mejorar en este desarrollo.
- **Ortografía, Redacción y Formato (10 puntos):** Se evalúa la precisión ortográfica, la claridad y coherencia en la redacción.
- **Referencias (10 puntos):** Usar referencias pertinentes siguiendo el formato APA 7, incluyendo literatura relevante y datos de expertos que aporten a la comprensión y análisis del problema.

La nota final se calcula con un total de 100 puntos con exigencia del 60 %. La extensión máxima permitida del informe es de 25 páginas, pero se permite utilizar Anexos para aquellas Figuras y Tablas que sean de tamaño considerable.

4. Observaciones

- La entrega debe ser subida al Campus Virtual. El plazo máximo se encontrará en el apartado de entrega.
- Las consultas deben ser realizadas al correo marcelo.alvarez.e@usach.cl.
- Para aprobar, el promedio de nota de los laboratorios debe ser $\geq 4,0$.
- El laboratorio debe ser realizado en duplas, a excepción de los estudiantes de doctorado, quienes deben realizarlo de manera individual.
- Enviar únicamente el informe en formato PDF con el formato `grupoX_labY.pdf`, donde *X* es el número de grupo, y *Y* el número de laboratorio. *Ejemplo: grupo3_lab1.pdf*.
- Se penalizarán las siguientes situaciones:

- La copia en el laboratorio será sancionada con la nota mínima.
- Se descontará puntaje por utilizar IA para generar, copiar y pegar texto sin siquiera modificarlo.
- Por cada hora de atraso en la entrega de un laboratorio, se descuentan 5 puntos.