



Profesor: Dr. Max Chacón Pacheco

Ayudante: [Miguel Salinas González](#)

Objetivos

- Extraer el conocimiento del problema asignado, mediante el uso del software R, utilizando el algoritmo de clustering K-means y realizar el análisis respectivo.
- Comparar los resultados con lo expuesto en la literatura encontrada y ver si se sustenta el conocimiento obtenido.
- Analizar por grupo e identificar aquellas características más relevantes, si clasifica mejor a una clase que otra e inferir conocimiento respecto a ello.

Instrucciones

1. El trabajo es en parejas.
2. Fecha de entrega: **Lunes 6 de noviembre del 2023 hasta las 23:59.**
3. La entrega consta de un informe de laboratorio (en PDF) y el código fuente con que se hicieron las pruebas.
4. La entrega tanto del informe (en PDF) como del programa debe ser en un archivo comprimido al correo miguel.salinas@usach.cl
5. Cualquier copia detectada entre los trabajos será calificada con nota mínima.

Aspectos importantes a considerar

Para obtener los resultados y cumplir los objetivos del laboratorio, se debe tener en cuenta los siguientes puntos:

Para trabajar con R deben utilizar el paquete Cluster y utilizar la función pam u otro paquete a conveniencia.

Deben decidir y justificar la normalización de las variables.

Deben justificar el criterio de proximidad a usar con el algoritmo.

UNIVERSIDAD DE SANTIAGO DE CHILE
FACULTAD DE INGENIERÍA
ANÁLISIS DE DATOS
LABORATORIO 2: AGRUPAMIENTO



Informe

El informe se debe regir por el reglamento de titulación v 1.3, apéndice C, apartado C.3 y contenerlos siguientes puntos:

Informe	Puntos a evaluar	Porcentaje
	Presentación, ortografía y redacción.	2%
	Introducción.	3%
	Marco Teórico: Clustering, algoritmo K-means y distancias utilizadas (máximo 2 páginas)	15%
	Pre-procesamiento: Se deben definir criterios para eliminar registros o columnas que presenten datos perdidos, outliers o que no aporten información relevante para el estudio del problema. En caso de normalizar datos fundamentar la decisión.	15%
	Obtención del Clúster: Variar parámetros de la función a utilizar de manera que se genere un clúster adecuado en base a las métricas de eficiencia en la clasificación, justifique la utilización de aquellos parámetros, además del criterio para seleccionar el clúster más adecuado. Decida y fundamente el criterio de proximidad de acuerdo a los datos.	10%
	Análisis de los resultados: Analizar el clúster e identificar aquellas características que sean más interesantes, entregando su significado en el dominio del problema y contrastar esta información con lo expuesto en la literatura	30%
	Respecto a los resultados obtenidos, el desarrollo del laboratorio y el método utilizado. Menciona aspectos positivos y a mejorar en este desarrollo.	20%
	Referencias: usar formato APA 6	5%

Observaciones

Consultas al mail: miguel.salinas@usach.cl

Para la entrega del laboratorio se debe enviar el archivo comprimido al correo indicado, hasta las 23:55 horas del día Lunes 6 de noviembre. En caso que no exista entrega hasta antes de ésta hora, se someterá el trabajo a descuento. Se descontará 1 punto por cada día de atraso.

La información de las bases de datos se encuentra en la página: <http://archive.ics.uci.edu/ml/>

UNIVERSIDAD DE SANTIAGO DE CHILE
FACULTAD DE INGENIERÍA
ANÁLISIS DE DATOS
LABORATORIO 2: AGRUPAMIENTO



Para aprobar el laboratorio es obligación realizar TODAS las experiencias