



Universidad de Santiago de Chile
Facultad de Ingeniería
Departamento de Ingeniería Informática

Asignatura : Análisis de Datos e Inteligencia de Negocios
Programa : Ingeniería Civil Informática
Ingeniería de Ejecución en computación e informática
Profesor : Dr. Felipe-Andrés Bello Robles
Ayudante : NA

Fecha Entrega : 23 de mayo de 2019.

TALLER 2: Análisis de Agrupamientos

Objetivos:

- Extraer el conocimiento del problema asignado, obteniendo mediante el 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.

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

Estructura del Trabajo	Puntos a evaluar	Porcentaje
	Presentación, ortografía y redacción, Referencias en formato APA 6	5%
	Introducción.	5%
	Marco Teórico: <ul style="list-style-type: none">• Clustering• Algoritmo K-means• Distancias utilizadas• Métricas de calidad	10%
	Pre procesamiento: Se debe realizar un pre-procesamiento de datos, donde se defina el criterio para eliminar registro o columna para datos perdidos, outliers o cualquier comportamiento anormal. Decidir y fundamentar en qué casos hay que normalizar las variables	10%
	Obtención del Clúster: Variar los parámetros del programa de modo de generar un clúster adecuado en base a las métricas de eficiencia en la clasificación (incluir el valor de los parámetros utilizados y explicar que significan y justificar el criterio del mejor clúster seleccionado). Decidir y fundamentar el criterio de proximidad en base a los datos.	15%

	<p>Análisis de los resultados: Analizar el clúster e identificar las características interesantes que se pueden obtener de éste, entregando el significado bajo el dominio del problema y contrastar con las reglas obtenidas en el laboratorio anterior y con la literatura.</p>	25%
	<p>Conclusiones: Respecto a los resultados obtenidos con el problema, al desarrollo del laboratorio, al método utilizado. Menciona aspectos positivos y a mejorar de este desarrollo.</p>	20%
	Código Fuente explicado o con comentarios.	10%

Observaciones:

- Consultas al mail felipe.bello@usach.cl
- El trabajo debe ser subido al Moodle del curso el día 23 de mayo de 2019 hasta las 11:55.
- Se debe utilizar la misma base de datos que utilizó en el taller 1, la información pertinente se encuentra en la página: <http://archive.ics.uci.edu/ml/>
- Para trabajar con R deben descargar el package “Cluster” y utilizar la función “pam”.
<http://cran.r-project.org/web/packages/cluster/cluster.pdf>
- Deben decidir y justificar la normalización de las variables.
- Deben justificar el criterio de proximidad a usar con el algoritmo.
- Para aprobar el laboratorio es obligación realizar TODAS las experiencias.