



Enunciado Laboratorio N°2: Inteligencia Computacional

Primer Semestre 2020

Profesor: Max Chacón Pacheco
Ayudante: Felipe-Andrés Bello Robles
Fecha Entrega: 1 de junio de 2020.

Objetivo:

Extraer el conocimiento del problema asignado, obteniendo mediante el software R, utilizando el algoritmo de clustering K-means y realizar el análisis respectivo. Luego debe comparar con los resultados revisados en la literatura encontrada y ver si se sustenta el conocimiento obtenido. Para ello se debe identificar el número de grupos a generar (en base a un criterio que usted defina). Luego se debe realizar el análisis (centroides) e identificar qué características podrían estar asociadas a una u otra clase.

Informe:

Se debe regir por el reglamento de titulación v 1.3, apéndice B, apartado B.3 y contener los siguientes puntos:

	Aspecto a Evaluar	Porcentaje
Informe	Presentación, ortografía y redacción.	10%
	Introducción. (1 página)	
	Marco Teórico. Clustering, K-means, distancias (2 páginas máx.)	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	15%
	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.	10%
	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 la literatura.	30%
	Conclusiones. (2 página máx). Respecto a los resultados obtenidos con el problema, al laboratorio, al método.	20%
	ANEXO: código fuente (R)	5%

Observaciones:

- Consultas al mail Felipe.bello@usach.cl
- Se descontará 1 punto por cada día de atraso
- Para trabajar con **R** deben descargar el package “**Cluster**” <http://cran.r-project.org/web/packages/cluster/cluster.pdf> y utilizar la función “**pam**”.
- 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