



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

Asignatura : Taller de minería de datos avanzada
Programa : Magister en Ingeniería Informática
Profesor : Dr. Max Chacón Pacheco
Ayudante : Felipe-Andrés Bello Robles

Fecha Entrega Oral : 23 de Agosto del 2017

Fecha Entrega Escrito : 30 de Agosto de 2017

TALLER 1: Planteamiento del problema y resolución mediante métodos de agrupamientos basados en modelos

Objetivos:

- Definir el problema a resolver mediante el conjunto de datos seleccionado
- Estudiar e interpretar los datos correspondientes a cada base de datos.
 - Para ello es necesario explicar de forma detallada el significado de clases, atributos y sus instancias, lo que permitirá obtener el correcto análisis del problema planteado.
- Brindar una solución mediante el uso de métodos de agrupamientos basados en modelos (Mcluster)

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

- Buscar la descripción de los datos en la página indicada en la parte de observaciones y así lograr el estudio correcto del problema., donde **deben** consultar los diferentes *papers* relacionados, los cuales han sido seleccionados porque utilizan dicha base de datos. De ésta manera obtendrá una interiorización apta para poder desarrollar los sucesivos laboratorios.
- Para el estudio de los atributos de su problema, se solicita además de la descripción detallada de éstos, aplicar técnicas de estadística descriptiva e inferencial según corresponda, tales como: Medidas de centralización (media, moda y mediana), distribución de probabilidades y Medidas de dispersión (rango y varianza), tests de hipótesis, análisis de varianza (ANOVA), etc...
- Para ello, es necesario utilizar "R" <http://www.r-project.org/> y así obtener un conocimiento un poco más profundo de la base de datos, por ejemplo las relaciones que existen entre las variables del problema mediante visualización.

- Finalmente, si es que es posible, se recomienda incluir información relevante entregada por un experto en el área de su problema, que ayude determinar la importancia de los atributos y descripción del problema.

Archivos de datos: generalmente archivo *.data, archivo *.names

Escrito:

Se debe elaborar un *paper* de máximo 6 páginas, según el formato:

<https://www.springer.com/gp/computer-science/lncs/conference-proceedings-guidelines>

Estructura del paper	Puntos a evaluar	Porcentaje
	Presentación, ortografía y redacción	5%
	Abstract e Introducción	10%
	Métodos y Datos: referencia a anexo de análisis estadístico	10%
	Resultados	15%
	Discusión	20%
	Conclusiones	20%
	ANEXO (máx. 3 paginas)	20%
	Análisis Estadístico: <ul style="list-style-type: none"> • Estadística Descriptiva • Estadística Inferencial • Conclusiones sobre la muestra en base al análisis estadístico y en relación a la resolución del problema. 	

Observaciones:

Consultas al mail Felipe.bello@gmail.com, Felipe.bello@usach.cl

El trabajo debe ser presentado de forma oral (50%) y escrita (informe 50%) en horario de clases el día 23 y 30 de Agosto de 2017. Disponen de 15-20 minutos de exposición y 10 para contestar preguntas de la comisión.

La información de las bases de datos se encuentra en la página:

<http://archive.ics.uci.edu/ml/>

Usar R package Mclust: <https://cran.r-project.org/web/packages/mclust/mclust.pdf>

Nota Final: Promedio simple de las experiencias.