

Tarea 4: Técnicas de Agrupación

Fernando Crema

June 16, 2015

Objetivo

Abastos Crema debe hacer análisis sobre muestras de laboratorio para encontrar agrupaciones. Usando los métodos **hcluster** o **kmeans** realice un informe en **.Rmd** respondiendo todas las preguntas presentadas a continuación justificando con un segmento de código de R (el dueño de la empresa quiere corroborar sus resultados) cada una de ellas exceptuando las mismas que comiencen con **Análisis** (en estas no necesita insertar un código).

Consideraciones de forma

- Recuerde que su **.Rmd** debe ser reproducible en cualquier máquina.

Consideraciones de contenido

- Toda pregunta es respondida usando solamente funcionalidades del paquete **stats** o funciones de los paquetes básicos de **R**.

Preguntas

- Se le presentan 5 archivos de entrada con 3 variables y variedad de filas. La 3ra variable para todos los dataset representa la columna de clase. Sin embargo, esta solo será usada con fines de analizar los resultados (generar las matrices de confusión).
- Para cada dataset:
 1. Graficar los puntos colocandole un color por clases a cada punto.
 2. **Análisis:** Decidir la métrica para obtener la matriz de disimilaridad y explique porqué usó la misma.
 3. Explorar los distintos métodos para agrupación jerárquica (en **hclust**).
 - Grafique cada dendrograma encontrado.
 - ****Análisis**:** Escoja el mejor dado el dataset.
 - Para cada método, imprima la matriz de confusión y analice la misma.
 4. Asumiendo el dendrograma “ganador”. Con qué **h** se cortaría el mismo si quisiera 2, 3, 4 o 5 clusters?
 5. **Análisis:** Escoger un **h** apropiado para cortar el árbol
 6. Graficar el dendrograma “cortado” en la altura escogida en el punto anterior.
 7. Escoger un **k** apropiado para usar el método k-medias.
 8. Grafique los clusters y sus centros.
 9. **Análisis:** Qué metodo funciona mejor? k-medias o jerárquico? Explique su respuesta.
 10. **Análisis:** Si obtenemos una instancia **I** nueva con $I = (3, 3)$.
 - En qué cluster estaría **I** asumiendo agrupación jerárquica por aglomeración?
 - En qué cluster estaría **I** asumiendo k-medias?

Paquetes de R

1. [RMarkdown](#), Documentos dinámicos para R.
2. [Agrupación Jerárquica I](#) Hierarchical Clustering I.
3. [Hierarchical Clustering II](#) Hierarchical Clustering II.
4. [Hierarchical Clustering III](#) Hierarchical Clustering III.
5. [Hierarchical Clustering IV](#) Hierarchical Clustering IV.
6. [Stats](#) Paquete Stats de R.

Consideraciones de entrega

1. La tarea es **estrictamente individual**.
2. Fecha de entrega: Viernes 07/08/2015, **Deadline: 11:59 pm**.