

1. Con el conjunto de datos “*data\_cluster.mat*”, emplee agrupamiento jerárquico para identificar el número de clusters que se ajustan al problema:
  - Use mínimo tres métodos de agrupamiento jerárquico (linkage simple, linkage completo, Ward, centroide, promedio entre grupos).
  - Apoyándose en el dendrograma, compare las agrupaciones realizadas para cada caso y realice un histograma que le permita evaluar aquellas agrupaciones comunes.
  - Establezca un criterio que le permita solucionar el problema. Explique y justifique su respuesta.
  - Concluya
2. Para el desarrollo de este taller utilice los datos tomados del repositorio de la UCI (<http://archive.ics.uci.edu/ml/datasets/Anuran+Calls+%28MFCCs%29#>), la cual es una base de datos de un problema multiclase para la clasificación de sonidos de ranas, los cuales se encuentran especificados por familia, género y especie.

El problema de agrupamiento será resuelto por usted, de acuerdo a la siguiente asignación:

<b><i>Tipo de agrupación</i></b>	<b><i>Último dígito de código estudiantil</i></b>
Familia y Genero	• Par
Género y Especie	• Impar

Tabla 1 Asignación de problema de trabajo

Esquema no Supervisado (use todos los datos):

- Diseñe dos modelos de agrupamiento jerárquico. Sin embargo, puede usar agrupamiento no jerárquico (K-means o Mediodes) de forma preliminar y combinarlo con agrupamiento jerárquico.
- Evalúe la calidad de sus grupos de acuerdo a la distancia entre ellos, SSE o Coeficiente de Silhouette.

Esquema Semi-Supervisado (5% de los datos para definir etiquetas, 10% para validar, el resto para entrenar )

- Etiquete la información con respecto al tipo de agrupación que le correspondió, como clases de salida del problema.
- Realice validación cruzada y estime el error de generalización para cada modelo.
- Analice sus resultados y describa los procedimientos en un informe.