

Pauta tarea 4

Introducción:

- Explicar el problema a resolver y las herramientas (clusterings) que se usan para resolverlo.
- Enumeración y explicación de las secciones que siguen

Marco teórico (incluir figuras / ecuaciones cuando corresponda):

- Explicar brevemente coeficientes cepstrales en frecuencias de Mel
- Explicar kmeans y kmeans++
- Explicar DBSCAN
- Explicar clustering aglomerativo
- Explicar métricas: completeness, homogeneity, v-measure, silhouette

Desarrollo:

- Implementar lectura dataset
- Modificar bench_k_means() para imprimir número de clusters
- Crear bench_DBSCAN():
 - o Indicar manejo inercia_
 - o Implementar datos no asociados v/s cluster extra
 - o Manejo de fallos en clustering
- Crear bench_agglomerative_clustering()
- Pruebas con 9 variantes sin PCA
- Pruebas con 9 variantes con PCA
- Comparar algoritmos con cada métrica, elegir peor y mejor método por métrica
- Analizar casos en que DBSCAN falla
- Analizar número de clusters de DBSCAN v/s métricas
- Analizar efecto de considerar outliers de DBSCAN como clase extra

Conclusiones

- Análisis general de resultados obtenidos en el desarrollo, indicar hasta qué punto se pudo resolver el problema, indicar cuál fue la mejor solución

Nota 1: Todos los elementos indicados arriba deben ser descritos en el informe.

Nota 2: En los ítems en que haya código asociado, se debe adjuntar la parte relevante del código.