

Proyecto #2: Clustering

Prof. Cristian López Del Alamo

En este proyecto usted cuenta con una base de datos de imágenes de rostros (*anger, contempt, disgust, fear, happy, sadness y surprise*). Su grupo deberá transformar las imágenes en vectores característicos (se le proporciona una librería de python para trabajar con wavelets, pero su equipo es libre de investigar una mejor). Una vez que las imágenes son transformadas en puntos k dimensionales, su equipo aplicará técnicas de *clustering* en este espacio. Se espera que cada *cluster* cuente con la mayor cantidad de imágenes con la misma expresión. Usted colocará como etiqueta del *cluster* el nombre de la expresión con mayor frecuencia en cada grupo.

Actividades

1. Utilizar el dataset **Base de Datos de Emociones**
2. Obtener los vectores característicos de cada imagen utilizando el haar wavelet. **Library en python**
3. Implementar los algoritmos GMM, K-Means, DBSCAN.
4. Utilizar un método de reducción de la dimensión como SVD o *principal component analysis*, además de fundamentar la dimensión reducida que se tomó para los experimentos. Para la reducción de la dimensionalidad de los datos puede usar librerías.
5. Evaluar la calidad de cada *cluster* verificando si los elementos en cada grupo, efectivamente pertenecen al mismo tipo de emoción.
6. Visualizar cada *cluster* colocando el nombre de la emoción más frecuencia en cada grupo.
7. Comparar los resultados entre los diferentes tipos de técnicas de *clustering*

NOTA: Cada equipo deberá subir un solo documento con la siguiente estructura:

1. Introducción.
2. Explicación de los métodos *clustering* así como las consideraciones tomadas en cuenta para generar los vectores característicos.
3. Experimentos. Es la parte más importante del proyecto y debe ser realizado de manera exhaustiva. Utilice gráficos y tablas para mostrar sus hallazgos.
4. Conclusiones: Redacte las conclusiones de acuerdo a los resultados.

Finalmente, anexar el link del github o el colab para verificar el código fuente.

NOTA: Deberá colocar el porcentaje de participación de cada estudiante al lado de su nombre entre paréntesis.

Ejemplo: Juan Paredes Melendes (10 %)