



**UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE INGENIERIA
DEPARTAMENTO DE ELECTRONICA
FUNDAMENTOS DE INTELIGENCIA COMPUTACIONAL
TRABAJO 3 – AGRUPAMIENTO**

Fecha de entrega: 27 de abril 2023

METODOLOGÍA

El siguiente trabajo debe ser resuelto en los grupos establecidos para realizar el proyecto de curso. El informe debe ser enviado al correo electrónico antes del 27 de abril a las 6am.

La entrega debe ser puntual, a la hora mencionada, si no se entrega a esa hora el trabajo será calificado sobre 4 y si se entrega después de 2 días de la fecha de entrega, es decir a partir del 29 de abril, se calificará sobre 3.

Para resolver el problema, puede usar un código o toolbox que ya tenga implementado los algoritmos de agrupamiento. Es importante que entienda bien el código usado y como se hace la variación de parámetros en cada algoritmo, que sepa explicar completamente el código usado y la forma como seleccionó los parámetros.

Para estimar el error con los datos de prueba deberá implementar un pequeño código de tal manera que pueda usar con nuevos datos (datos de prueba) el clasificador obtenido al aplicar cada algoritmo de agrupamiento.

Vía correo electrónico, el día 27 de abril, debe enviar los archivos que hayan creado y el informe.

NOTA: Para la solución de esta tarea SI pueden utilizar toolbox o código realizado por otra persona, sin embargo, para estimar los errores de prueba seguramente debe programar el clasificador que obtenga (centros encontrados, fórmulas de distancias y fórmula de grado de pertenencia) para poder estimar la salida con nuevos datos sin volver a hacer el proceso de identificación de centros.

EVALUACION

El 50% de la nota corresponde al informe que entregan por escrito, el otro 50% corresponde a la sustentación. Cada uno de los puntos del trabajo se evaluará de la siguiente manera: El punto 1 tiene un valor del 25% y el punto 2 tiene un valor del 25%, sumando así el 50% correspondiente al informe escrito. Bonus: El grupo que realice el punto 3 con porcentaje de error de prueba y de entrenamiento menores al 10% se le adiciona un 0.5 en el valor final de la nota de la tarea.

OBJETIVOS

- Afianzar los conceptos de algoritmos de agrupamiento y de clasificación.

PUNTOS

Para este trabajo se va a resolver el mismo problema abordado en la tarea de Redes Neuronales. Usar los mismos datos para entrenamiento y prueba que usó en la tarea 1. En cada punto usar los datos de Esterificación (parte A de cada punto) y los datos de Tiosulfato (parte B de cada punto)

1. Utilice el algoritmo K-Means para agrupar los datos de entrenamiento y encontrar los centros que permiten diferenciar entre las clases esperadas.

Verifique el desempeño del algoritmo con 4 valores diferentes de *clusters* (k), estime el error de entrenamiento. Recuerde que 2 o más *clusters* pueden corresponder a una de las clases esperadas.

Utilice la misma correspondencia de *clusters* a cada clase que asoció para los datos de entrenamiento. Estime el error con los datos de prueba.

Se espera una tabla donde se presente el error para los datos de entrenamiento y prueba con 4 valores de K diferentes.

2. Para los mismos datos elija si usa el algoritmo FC-Means o GK-Means, y repita el procedimiento del punto 1.

Se espera una tabla donde se presente el error para los datos de entrenamiento y prueba con 4 valores de K diferentes.

Revise los grados de pertenencia de cada dato de entrenamiento a cada *cluster* para verificar si tiene valor alto de pertenencia al *cluster* asignado o si hay alta incertidumbre, es decir si los grados de pertenencia de cada dato a todas las clases es similar.

Revise los grados de pertenencia de cada dato de prueba a cada *cluster* para verificar si tiene valor alto de pertenencia al *cluster* asignado o hay alta incertidumbre, es decir si los grados de pertenencia de cada dato a todas las clases es similar.

Qué relación existe entre la incertidumbre que tiene en la clasificación y el % de error en entrenamiento y prueba?

3. Opcional: Para los mismos datos aplique el algoritmo LAMDA, y repita el procedimiento del punto 1.

En ambos casos debe mostrar en tablas la forma en que eligió los parámetros propios de cada algoritmo (ejemplo: m , C , criterio de parada). En la tabla debe reportar el desempeño de cada algoritmo (porcentaje de error de entrenamiento y prueba) para cada prueba.

Compare los resultados obtenidos con los que usted alcanzó en la tarea de Redes Neuronales.