

**Universidad Interamericana PR – Bayamón**  
**Matemáticas Discretas para Ingeniería de Computadoras**  
**Proyecto Final**

---

## Objetivo

- Aplicar conceptos definidos en clase.
- implementación en un lenguaje de alto nivel de un algoritmo basado en grafos
- Análisis de los resultados de la implementación realizada

## Descripción

El proyecto consistirá en implementar el algoritmo llamado `InterCovid(d, k)` luego del cual se realizará un análisis teórico. El algoritmo a implementar se muestra en la Figura-1

### ***InterCovid (d,k)***

$Entrada = \begin{cases} d \rightarrow \text{datos sin agrupar} \\ k \rightarrow \text{cantid. de grupos} \end{cases} \quad Salida = \text{Prediccion, Exactitud}$

(1) Seleccionar la estrategia para inicializar los centros de los grupos

**Repetir:**

(2)       Actualizar los centros de los grupos usando la estrategia indicada

(3)       Calcular la distancia de los objetos(puntos) a los centros

(4)       Asignar cada objeto a un grupo basado en la mínima distancia

(5) **Hasta**   (ningún objeto se mueva de grupo o alcance las 100 iteraciones)

(6) Imprimir la exactitud

(7) Mostrar las gráficas de los grupos predichos

Figura-1: Pseudo- código del algoritmo `InterCovid(d, k)`

## Funcionamiento del algoritmo

InterCovid(d, k) agrupa los datos basándose en la similitud entre cada uno y buscando que los datos que están dentro de un grupo (clúster) sean los más similares entre sí. Ver ejemplo explicado por profesor.

## Fórmulas de distancias

Para el proyecto usaremos las métricas mostradas en la Tabla-1. En dicha tabla los puntos p y q están definidos como:

$p = (p_1, p_2, \dots, p_n)$  y  $q = (q_1, q_2, \dots, q_n)$  Si  $n=2$  entonces estamos en dos dimensiones que es lo que usualmente estamos acostumbrados a trabajar

Tabla-1: Métrica de distancias

Nombre	Formula
(1) Euclidiana	$d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}$ $= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}.$
(2) Manhattan	$d_1(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^n  p_i - q_i ;$

## Estrategia de los Centros

Para el manejo de los centros, en este proyecto se utilizarán los siguientes enfoques:

### 1. Inicialización

Asumiendo que X (datos en archivo de datos) es la matriz que contiene las observaciones. Se usarán dos estrategias:

- Aleatorio:** Selecciona k observaciones(puntos) de X al azar. Estos serán los centros
- Primeros elementos:** se escogen los primeros elementos del archivo de datos X como centros de los grupos.

## 2. Actualización de centros:

Usaremos dos opciones:

- Promedio aritmético:** El nuevo centro es hallado, promediando las coordenadas de los puntos.
- Mediana:** El nuevo centro es calculado hallando la mediana de las coordenadas de los puntos.

## Exactitud

La precisión es un criterio para medir la calidad de un agrupamiento. La definición formal es dado de la siguiente manera. Dado un conjunto de grupos  $G$  (predicciones) y un conjunto de clases  $C$  (verdaderas), ambos particionando un total de  $N$  puntos de datos, la precisión puede ser definida como:

$$\frac{1}{N} \sum_{g \in G} \max_{c \in C} |g \cap c| \quad \text{-----} \quad (1)$$

Para clarificar esta fórmula veamos un ejemplo. Sean tres grupos  $G1$ ,  $G2$ ,  $G3$  que ha resultado después de ejecutar el algoritmo, tal como se aprecia en la Figura-2 ( $N=17$ )

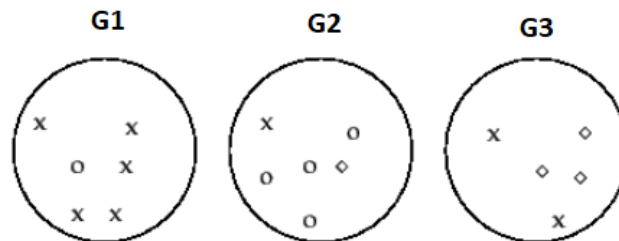


Figura-2 Ejemplo de la medida Precisión con 3 grupos

Asumimos que:

x = elementos que realmente pertenecen al G1

o = elementos que realmente pertenecen al G2

◇ = elementos que realmente pertenecen al G3

El cálculo de precisión sería:  $P = \frac{1}{17} (5 + 4 + 3) \approx 0.70$

## Requisitos del proyecto

- El programa debe tener un archivo “main” conteniendo el desarrollo de algoritmo y utilizando funciones escritas por usted.
- El programa recibe como input un archivo de texto el cual contendrá los datos. El archivo será dado por el profesor, así como el numero k
- Limite el máximo número de iteraciones a 100 (una vez llegado a 100 para y muestra los resultados hasta ese momento)
- Los resultados serán presentados como se muestra en la Tabla-2

**NOTA:** Para obtener todos los puntos del proyecto es necesario que creen sus propias funciones. Por ejemplo, crear una función para hallar las distancias, así como funciones propias para todas las estrategias indicadas (excepciones serán aceptadas para mostrar cualquier gráfica y crear números aleatorios)

Tabla-2: Resultados

<b>Distancia:</b>		
<b>Actualización:</b>		
<b>Inicialización:</b>		
<b>Nombre Archivo</b>	<b>K</b>	<b>Exactitud</b>
Data1	3	
Data2	3	
Data3	2	

# Informe

Su informe debe estar bien escrito en inglés o español. El informe contendrá los siguientes elementos:

- 1) Introducción: breve descripción del proyecto
- 2) Programa
  - a) Explique las pruebas y los resultados generados (muestre imágenes en pantalla de las salidas del programa)
  - b) Explique cada sección del programa (muestre imágenes en pantalla de parte del código explicando cómo funciona)
  - c) Usando una tabla, enumere todas las funciones creadas. Las columnas de la tabla serán: nombre, descripción, entrada, salida de la función.
- 3) Resultados
  - a) Mostrar los resultados según la Tabla-2 (mostrada arriba)
  - b) Analice la tabla e indique cual es el mejor resultado y porque cree que es el mejor
- 4) Complejidad de tiempo
  - a) Realice un análisis del tiempo que tarda su código para el peor caso. Exprese el resultado en notación "***Big O*** "
- 5) Conclusión:
  - a) Analice su diseño en términos de cómo los resultados cumplieron con sus expectativas.
  - b) Describa todos los problemas principales que encontró y como fue solucionado
  - c) Cosas que probaste y que no funcionaron.
  - d) Haz una lista de todas las cosas que aprendiste
- 6) Apéndices: Adjunte aquí todo el código fuente