

# Formulación Matemática de K-Means y Algoritmo Genético

Javier Horacio Pérez Ricárdez

marzo del 2025

## 1. Formulación Matemática de K-Means

El algoritmo K-Means es un método de agrupamiento (clustering) que busca minimizar la suma de las distancias al cuadrado entre los puntos de datos y los centroides de los clusters. La función objetivo que se minimiza es:

$$J = \sum_{i=1}^n \sum_{j=1}^k w_{ij} \cdot \|\mathbf{x}_i - \mathbf{c}_j\|^2$$

Donde:

- $n$  es el número de puntos de datos.
- $k$  es el número de clusters (en este caso,  $k = 1$ ).
- $\mathbf{x}_i$  es el vector de características del punto  $i$ .
- $\mathbf{c}_j$  es el centroide del cluster  $j$ .
- $w_{ij}$  es un indicador binario que vale 1 si el punto  $i$  pertenece al cluster  $j$ , y 0 en caso contrario.

En el código, se utiliza la PEA como peso para cada punto de datos. Por lo tanto, la función objetivo se modifica para incluir la PEA:

$$J = \sum_{i=1}^n \text{PEA}_i \cdot \|\mathbf{x}_i - \mathbf{c}_j\|^2$$

Donde:

- $\text{PEA}_i$  es la Población Económicamente Activa del municipio  $i$ .

## 2. Formulación Matemática del Algoritmo Genético

El Algoritmo Genético (GA) es un método de optimización basado en la selección natural. En este caso, se utiliza para minimizar una función objetivo que depende de la distancia ponderada por la PEA.

### Función Objetivo

La función objetivo que se minimiza es:

$$f(\mathbf{x}) = \sum_{i=1}^n \text{PEA}_i \cdot d(\mathbf{x}, \mathbf{x}_i)$$

Donde:

- $\mathbf{x} = (lat, lon)$  es la ubicación propuesta (individuo en el GA).
- $\mathbf{x}_i = (lat_i, lon_i)$  es la ubicación del municipio  $i$ .
- $d(\mathbf{x}, \mathbf{x}_i)$  es la distancia entre la ubicación propuesta y el municipio  $i$ , calculada usando la fórmula haversine.
- $\text{PEA}_i$  es la Población Económicamente Activa del municipio  $i$ .

## Operadores del Algoritmo Genético

- **Selección:** Se utiliza selección por torneo para elegir los individuos que pasarán a la siguiente generación.
- **Cruce:** Se utiliza cruce blend ( $BLX-\alpha$ ) para combinar dos individuos y generar descendencia.
- **Mutación:** Se aplica mutación gaussiana a los individuos, limitando los valores de latitud y longitud dentro de los rangos válidos.

## 3. Uso de la PEA (Población Económicamente Activa)

La PEA se utiliza como peso en ambas formulaciones (K-Means y Algoritmo Genético) para dar mayor importancia a los municipios con una mayor población económicamente activa. Esto asegura que las ubicaciones sugeridas estén más cerca de los municipios con mayor PEA, lo que es relevante para el problema de optimización de ubicaciones.

En el código, la PEA se incorpora en:

- La función objetivo del Algoritmo Genético, donde se multiplica por la distancia.
- La ponderación de los puntos en el algoritmo K-Means.

## 4. Fórmula Haversine

La distancia entre dos puntos geográficos se calcula usando la fórmula haversine:

$$d = 2R \cdot \arcsin \left( \sqrt{\sin^2 \left( \frac{\Delta\phi}{2} \right) + \cos(\phi_1) \cdot \cos(\phi_2) \cdot \sin^2 \left( \frac{\Delta\lambda}{2} \right)} \right)$$

Donde:

- $R$  es el radio de la Tierra (6371 km).
- $\phi_1, \phi_2$  son las latitudes de los dos puntos en radianes.
- $\lambda_1, \lambda_2$  son las longitudes de los dos puntos en radianes.
- $\Delta\phi = \phi_2 - \phi_1$  es la diferencia de latitudes.
- $\Delta\lambda = \lambda_2 - \lambda_1$  es la diferencia de longitudes.

## Resumen

- **K-Means:** Minimiza la suma de las distancias al cuadrado ponderadas por la PEA.
- **Algoritmo Genético:** Minimiza la suma de las distancias haversine ponderadas por la PEA.
- **PEA:** Se utiliza como peso para priorizar municipios con mayor población económicamente activa.

Este enfoque permite encontrar ubicaciones óptimas que minimicen la distancia total ponderada a los municipios, considerando su importancia económica.