

1. Cálculo de la distancia euclidiana

La función `calcular_distancia(a, b)` calcula la distancia entre dos puntos. Para ello, utiliza la fórmula de la distancia euclidiana, que consiste en calcular la raíz cuadrada de la suma de los cuadrados de las diferencias entre las posiciones de los puntos.

2. Inicialización de los centroides

La función `inicializar_centroides(X, metodo='random')` genera los centroides iniciales para el algoritmo K-Means. Existen dos métodos de inicialización:

- **Método 'random':** Se generan centroides aleatorios dentro del rango mínimo y máximo de los datos.
- **Método 'dataset':** Se seleccionan aleatoriamente 4 puntos del dataset como centroides iniciales.

3. Asignación de puntos a clusters

La función `asignar_cluster(X, centroides)` asigna cada punto del conjunto de datos al cluster cuyo centroide esté más cerca. Para ello, se calcula la distancia de cada punto a todos los centroides y se asigna el punto al cluster correspondiente al centroide con la menor distancia. En caso de empate se asigna el punto en cuestión al cluster con el menor número de puntos ya asignados, para ello se utiliza un contador de los puntos que se han asignado a cada cluster. Además, se han añadido comprobaciones para evitar detenciones inesperadas en la ejecución del código.

4. Actualización de centroides

La función `actualizar_centroides(X, clusters, k)` recalcula la posición de cada centroide como el promedio de todos los puntos que le fueron asignados en la iteración actual.

5. Implementación del algoritmo K-Means

La función `k_means(X, k=4, metodo='random', max_iter=100)` implementa el algoritmo completo. Primero, se inicializan los centroides y se copian para compararlos en futuras iteraciones. Luego, en cada iteración se realiza la asignación de puntos a los clusters y la actualización de los centroides. El proceso se repite hasta alcanzar el número máximo de iteraciones o hasta que los centroides dejen de cambiar (convergencia).

6. Generación de datos sintéticos

Se crean datos de prueba como se indica en la descripción de la práctica.

7. Visualización de los resultados

Finalmente, se grafican los puntos del dataset coloreándolos según el cluster al que fueron asignados. Los centroides se marcan con una "X" roja.