

# Implementación de Kmeans paralelizado

Joel Challco Tito  
Estudiante de Ciencia de la Computación  
Univesidad Católica San Pablo  
Arequipa, Perú  
Email: joel.challco@ucsp.edu.pe

## I. INTRODUCCIÓN

Kmeans es un método de agrupamiento, el cual tiene como objetivo el particionamiento de un conjunto N de datos en K grupos o clusters en el que cada dato pertenece a un cluster, este método es utilizado en minería de datos.

## II. ALGORITMO KMEANS

Kmeans paso a paso:

1. Especifique el número de grupos K.
2. Inicialice los centroides de forma aleatoria.
3. Asigna los datos al centroide mas cercano a cada punto.
4. Actualiza los centroides respecto al conjunto de datos.
5. repetir el paso 3 hasta que los centroides se actualicen minimamente.

## III. KMEANS VS KMEANS PARALELO

En este experimento se implemento el Kmeans secuencial y Kmeans paralelo. Las comparaciones se hicieron con un conjunto de datos en 2 dimensiones y 40000 puntos. Se obtuvieron los siguientes resultados.

- Resultado para el conjunto de datos con 2 cluster.

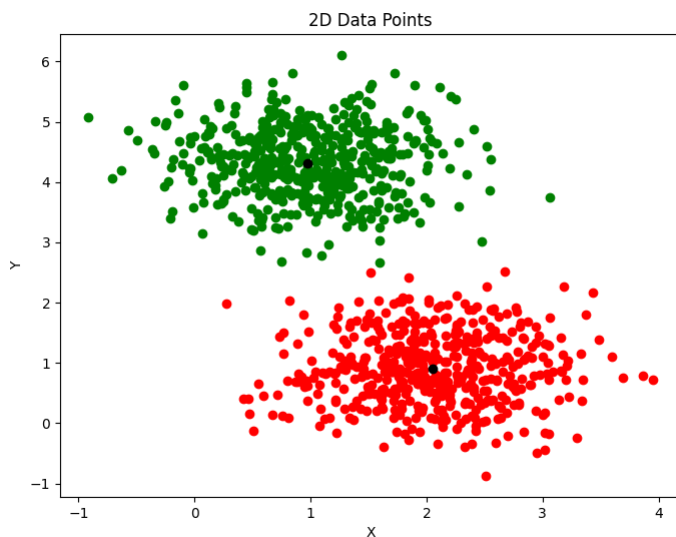


Figura 1. Kmeans, 2 cluster,40000 puntos

- Resultado para el conjunto de datos con 5 cluster.

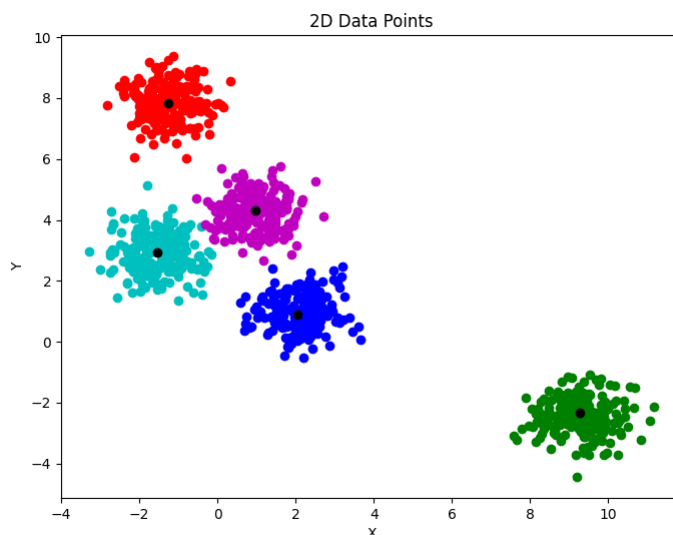


Figura 2. Kmeans, 5 cluster,40000 puntos

Para la paralelizacion del Kmeans, se dividió el conjunto de datos para poder calcular las distancias con cada centroide, de tal forma que a cada thread se le pueda asignar una cantidad de datos y no incurrir en una falsa paralelizacion.

## IV. CONCLUSIÓN

Se han realizado las pruebas con conjuntos de datos mas grandes, para ambos casos obteniendo resultados identicos. En cuanto a la velocidad tienen el mismo tiempo de ejecucion para a cantidad de puntos que tenemos, aplicando los mismos algoritmos a cantidades de puntos mayores como 100000 puntos se puede notar una diferencia mas notable en los tiempo.

Por lo tanto, decimos que la paralelizacion no es mejor para cantidad de datos pequeños.

## REFERENCIAS

savyakhosla, *K-means++ Algorithm*,  
<https://www.geeksforgeeks.org/ml-k-means-algorithm/>  
Yash Kothekekar, *Implementing K-Means Clustering Algorithm in C++ with an Example.*, <https://coderspacket.com/implementing-k-means-clustering-algorithm-in-c-with-an-example>