

Assignment 4 *

Ishaan Kapoor

February 21, 2022

1 Streaming Algorithms

A

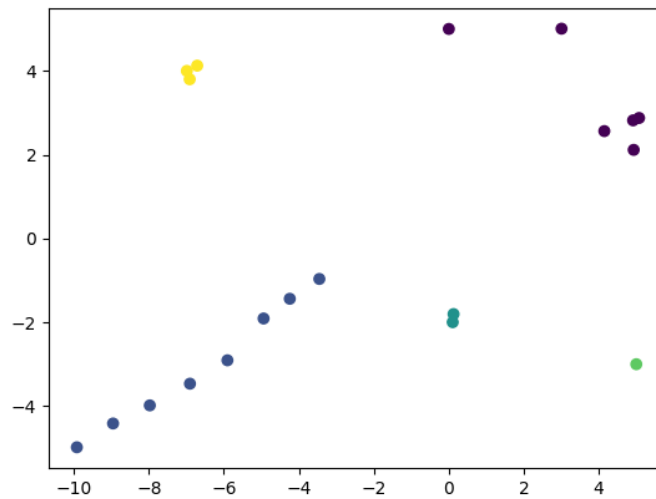


Figure 1: Single Link

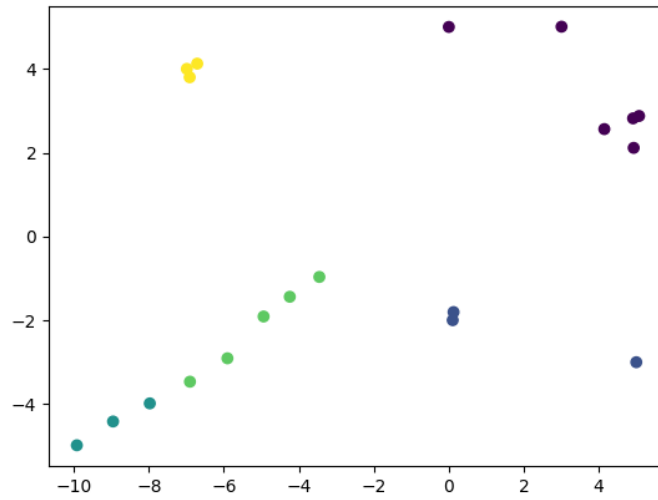


Figure 2: Complete Link

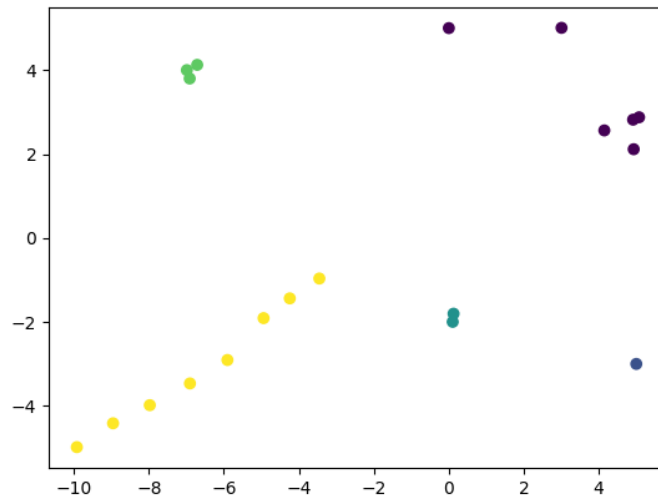


Figure 3: Mean Link

B

In my opinion single link does the best job as it produces the same results as mean link but has a time complexity of $O(n^2)$ as opposed to that of mean link which uses $O(n^3)$

2 Assignment-based Clustering

A

the 4-center cost $\max_{x \in X} d(x, C) = 26.4695$

the 4-means $\sqrt{\frac{1}{|X|} \sum_{x \in X} (d(x, C))^2} = 11.044$

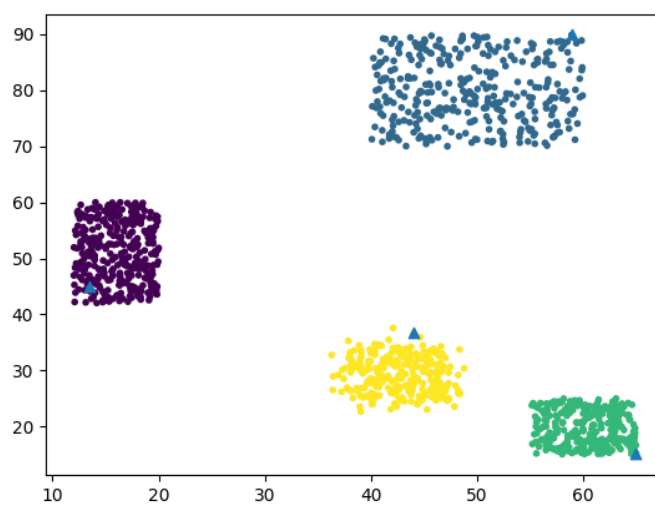


Figure 4: Gonzales Centers (marked with Δ)

B

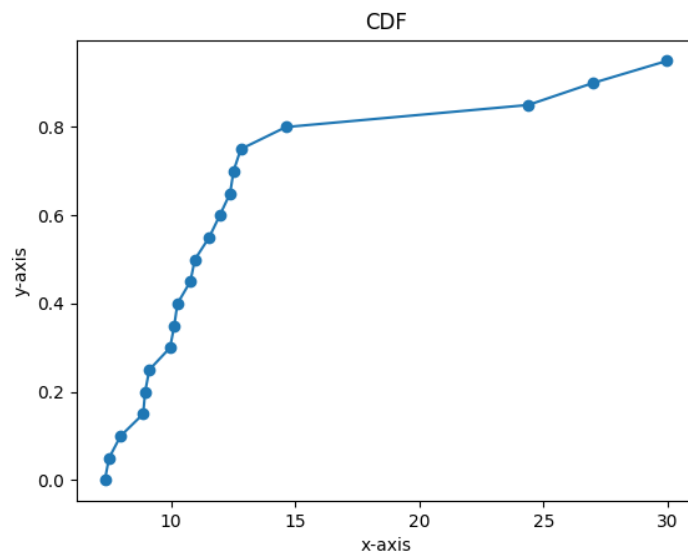


Figure 5: 4 means cost

C

a) 4 means cost: 8.554

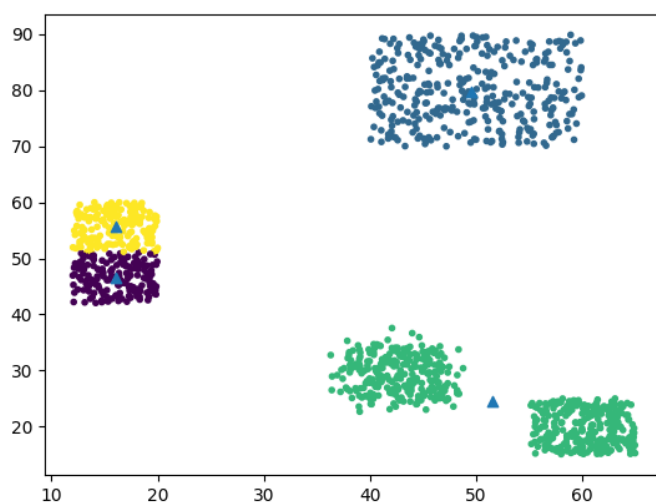


Figure 6: first 4 points as c

b) 4 means cost: 5.913

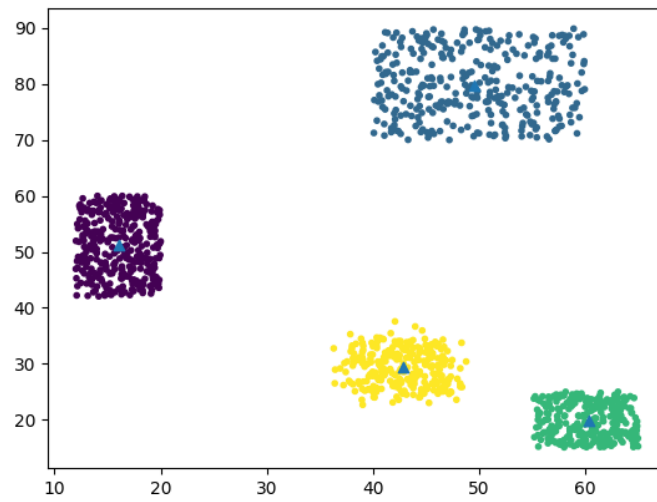


Figure 7: Gonzales Algorithm

c) ratio of centers that were the same: $\frac{4}{20}$

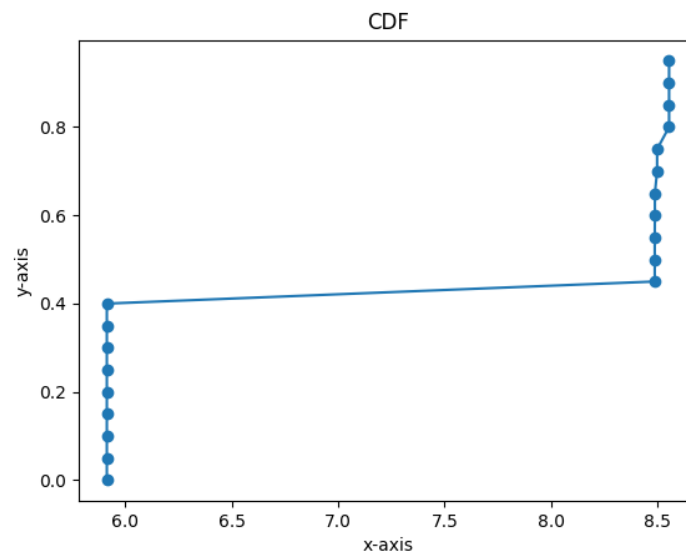


Figure 8: K means++ cost