

COMP 5630 Fall 2022 Assignment 7

Will Humphlett (wah0028)

November 8, 2022

1. K-Means Implementation

(a) Argue the exponential runtime

Given $k = 2$, data will be partitioned into 2 clusters, implying 2^{n-1} possible partitionings. This provides a runtime of $O(n * 2^{n-1})$.

2. Analyze the Effect of K

(a) Objective Plot

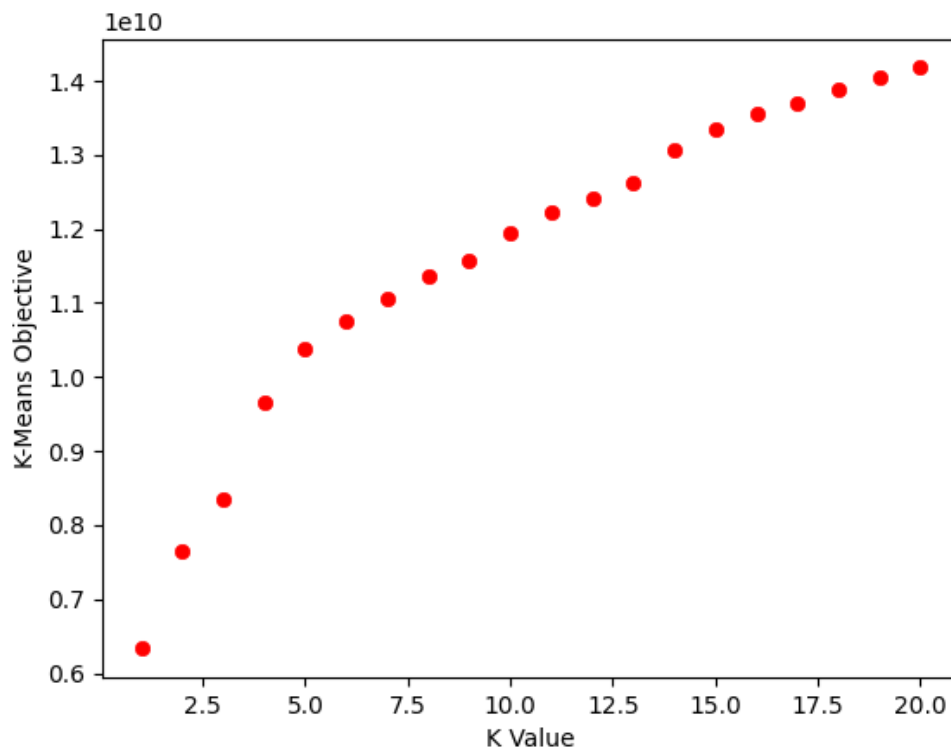


Figure 1: K-Means against K Value

(b) Find the elbow

The most obvious elbow in the graph above is at $k = 5$. However, this goes against initial impressions of looking at the data as the k-means objective improves by up to 50% as k value increases.

3. Analyze the Effect of Initialization

(a) $img = 5, k = 2$

Each image either has a silhouette of Mario or is only one color (likely the average color of the image). The silhouette implies 2 clusters and the single color implies equal centroids of each cluster or only one cluster.

(b) $img = 5, k = 10$

Each image ranges between 1 and 4 full colors/clusters. Each centroid is wavers slightly between each image.

(c) $img = 8, k = 2, 3, 4, 5$



Figure 2: $k=2$



Figure 3: $k=3$



Figure 4: $k=4$



Figure 5: $k=5$

(d) Propose Heuristic

Furthest point heuristic could work to help spread the initial placement of the centroids and produce better segmentation results.