Homework 3
Problem 1
Ayad Masud
733009045

I certify that I have personally done the coding, generated the figures and written the report without aid from anybody else, and that I have not plagiarized, self-plagiarized, or used AI-generated text. I certify that I have acknowledged any sources I used to complete this assignment. ARM.

1 Part 1: K-Means Clustering of Soccer Image

Figure 1 shows the soccer image reconstructed using K-means from values of k= 2 ... 10. We can see when k=2 only the most dominant colors are shown in the image, green and white. At k = 2 the shapes are not well defined, and we cannot really see the picture clearly. As we increase the value of k more and more color is added to the image. A trend we can see with increasing values of k is that more detail is preserved because more colors are used for the k-means in turn allowing us to perceive the entire image in full detail.

Figure 1. Image reconstruction of soccer image using K-means and vector quantization.

2 Part 2: K-Means Clustering of Holi Image

Figure 2 shows the Holi image reconstructed using K-means from values of k= 2 ... 10. We see a similar trend to part 1 with the soccer image. At k=2 we only see the 2 most dominant colors that are present in the image. At k=2 we cannot really see any details of the image. As we increase the value of k we get more and more color. This adds details to the image and lets us so the complete image in focus unlike at lower values of k where the number of colors is too small to preserve detail in the image.

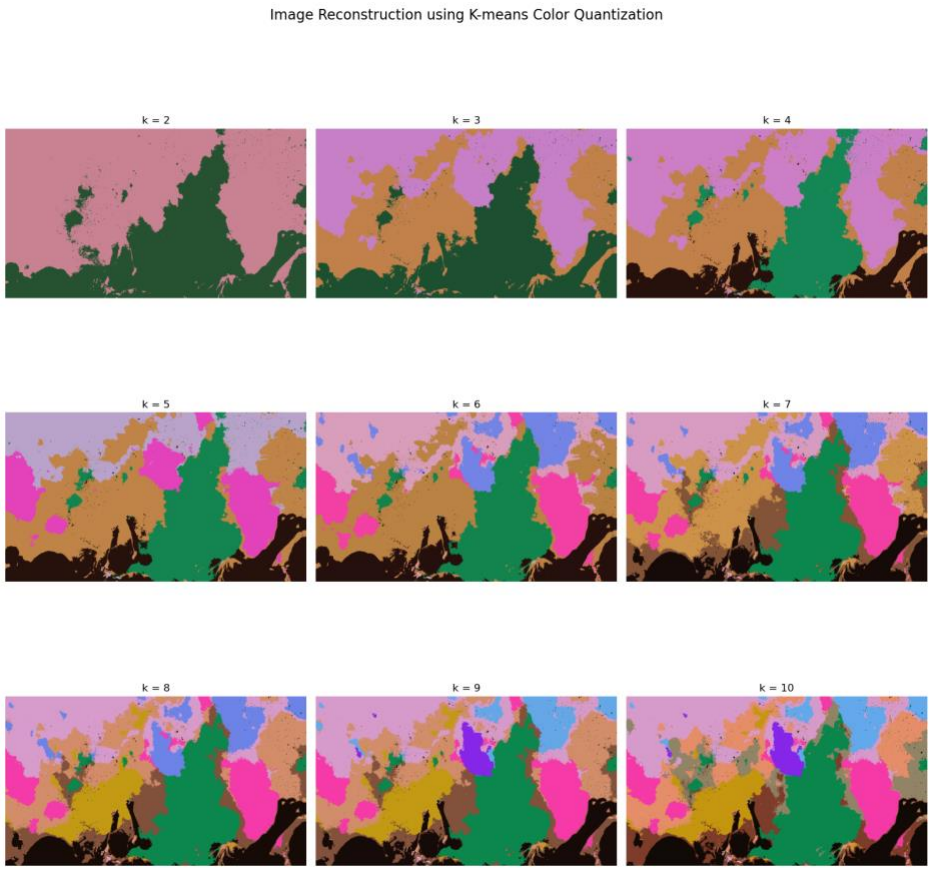

Figure 2. Image reconstruction of Holi image using K-means and vector quantization.

3 Part 3: SSE between Original and Reconstructed Image

Figure 3 plots the SSE between the original image and the reconstructed image for both the soccer image and Holi image.

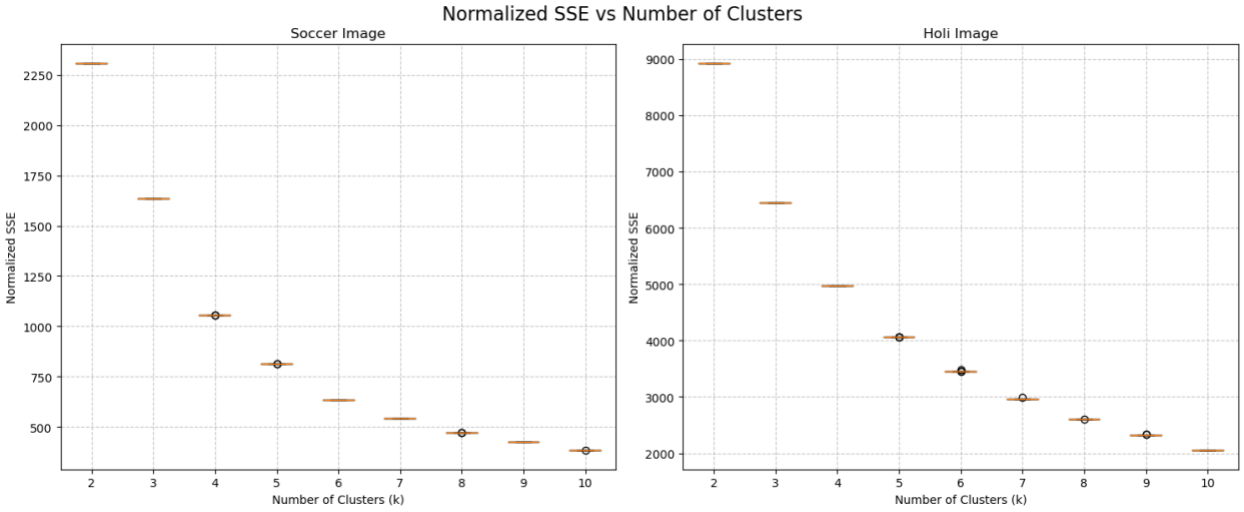

Figure 3. Plot between the SSE of the original image and the reconstructed image for both the soccer and Holi images. A trend we see in both graphs is that the normalized SSE decreases as k increases. This is because with more clusters and higher values of k, we can represent the original image more completely.

These graphs don't have a clear elbow point or a point where we can clearly tell that is the optimal value of k. On these graphs we have a much larger range to pick from for the most optimal value of k. The variation in SSE on different runs decreases as we run the algorithm over and over. This shows that the algorithm is more stable at higher levels of k.

4 Resources used to achieve this goal

Canvas: Homework template

Python Libraries: NumPy, pandas, matplotlib, sci-kit learn

5 References

scikit-learn. "Sklearn.cluster.KMeans — Scikit-Learn 0.21.3 Documentation." *Scikit-Learn.org*, 2019, scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html.