

# **CS 6375: MACHINE LEARNING**

## **ASSIGNMENT 5**

DHARA PATEL(DXP190051)

### **K-means Clustering for Image Compression**

The report contains the following:

1. The size comparison and evaluation for both images (Koala.jpg and Penguins.jpg) before and after compression
2. The outputs (images) that were generated after data compression using K-means for Clustering for both, Koala.jpg and Penguins.jpg.
3. The answers to the two questions asked in the assignment.

**Koala.jpg:** The original image size is **762.53 KB** & the number of Iterations used for k-means are 50.

K value	Size after Compression (in KB)	Ratio
2	129.19 KB	5.902
5	168.74 KB	4.519
10	172.53 KB	4.420
15	160.43 KB	4.753
20	160.64 KB	6.747

The image below is the **original** Koala image (without compression)



The images below are the outputs of **data compression** using different k values for Koala.jpg.

K = 2



K = 5



K = 10



K = 15



K = 20



**Penguins.jpg:** The original image size is **759.6 KB** & the number of iterations used for k-means are 50.

K value	Size after Compression (in KB)	Ratio
2	83.22 KB	9.128
5	100.66 KB	7.546
10	115.59 KB	6.572
15	111.79 KB	6.560
20	111.79 KB	6.560

The image below is the **original** Penguins image (without compression)



The images below are the outputs of **data compression** using different k values for Penguins.jpg.

K = 2



K = 5



K = 10



K = 15



K = 20



## 1. Is there a tradeoff between image quality and degree of compression?

Yes, there is a tradeoff between the two.

'K' represents the degree of compression here, so the smaller the value of K, the fewer the clusters we get and thus, less colors to represent the image.

From the above explanation and the compressed images, we can infer that a lower quality image will be generated for smaller 'K' values as lot of details in the image are compromised while compression. Whereas a better-quality image will be generated for higher 'K' values as it can show more colors due to a larger number of clusters. However, it takes a longer time to execute image compression with higher values of K.

## 2. What would be a good value of K for each of the two images?

The K values 15 and 20 generates compressed images which are decent for the given images, Koala.jpg and Penguins.jpg. The output image file shows the reasonable compression ratio with most colors which are identifiable.

For Koala.jpg, we get a compression ratio of 4.42 even with  $K = 10$ , which is quite close to that for  $k=15$  and 20. Keeping in mind the time for execution for image compression, 10 can be a good choice for k as it has almost the same compression ratio as of greater values of k.

For Penguins.jpg, we get a compression ratio of 6.56 with  $K = 15$  and  $k = 20$  both. Keeping in mind the time for execution for image compression, 15 can be a good choice for k.