Project Report

Nikhil Pareek (nxp170014)

Machine Learning CS 6375.001 S19

There is one code available, k\_means.py which uses k-means clustering algorithm for Image Compression technique. Further two images have been provided Koala.png and Penguins.jpg to test the written code for Image compression using k-means clustering. The results of both the images for different k values have also been submitted in the folders Koala Images and Penguins Images. The respective values of K means values that have been used are 2, 5, 10, 15 and 20. TAnd also the respective images developed from image compression for each k value is displayed while the code is run.

**Answer1:** Yes there is an inverse relationship between the degree of compression (k) and the quality of the image. This means that as the degree of compression increases (as the k values inccreases, the quality of the image decreses subsequently. The same could be seen in the follwing images for the Koala.jpg:

1) Figure 1: K = 2 (Just two colors to explain the Koala image)



2) Figure 2: K = 5 (Just five colors to explain the Koala image)



3) Figure 3: K = 10 (Just ten colors to explain the Koala image)



4) Figure 4: K = 15 (Just fifteen colors to explain the Koala image)



5) Figure 5: K = 20 (Just twenty colors to explain the Koala image)



**Similar observations could be made for the second image that is Penguins.jpg**

**Answer2:** A good value of k for the first image, which is Koala image would be k =15 since the image reproduces an appropriate image with a size of only 390.4 Kb (half of the size of the original image 780 Kb). Reference Table 1.

And in the second image which is Penguins Image, a suitable value of K would be 20 with a size of 277.7 Kb (one fourth of the size of the original image which is 777.8 Kb). Hence in both the cases, these values are sufficient for data compression and the same time successfully reproduce a very good quality image back. Reference Table 2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Original Image** | **K = 2** | **K = 5** | **K = 10** | **K = 15** | **K = 20** |
| **Size of the image(Kb)** | 780 | 65 | 166 | 304.4 | 390.4 | 478.5 |

Table:1 (Koala Image)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Original Image** | **K = 2** | **K = 5** | **K = 10** | **K = 15** | **K = 20** |
| **Size of the image(Kb)** | 777.8 | 35.1 | 77.8 | 166 | 222.3 | 277.7 |

Table:2 (Penguins Image)

How to Run the k\_means.py from command terminal interface:

>> python k\_means.py <argument1>

<argyment1> is the name of teh image that is to be compressed and is present in your working directory

Dependies:

The code was developed on Python 3.7.1, using Sypder IDE and a list of dependies are as follows:

1) import sys

2) from matplotlib import pyplot as io

3) import numpy as np

4) from PIL import Image

5) import random as random

6) from tqdm import tqdm

Kindly note:The k means clustering code would typically take around 4-5 minutes to run completely and display the corresponding images, for the given values of parameter k.

Kindly note: The code was developed on Python 3.7.1, using Sypder IDE.

Email id: **nxp170014@utdallas.edu**