E0-270: Assignment 2

Due date: April 12, 2023

Data

You are given an RGB image of shape $512 \times 512 \times 3$. Each pixel of the image is considered to be a data point.

Task: K-Means Clustering

The task is to cluster the pixels of the image using the K-Means algorithm, with given number of clusters $k = 2, 5, 10, 20, 50$.

For the algorithm, you can refer to Section 9.1 in the book "Pattern Recognition and Machine Learning" by Bishop.

You are given a folder containing the image and three python files 'main.py', 'model.py', and 'utils.py'.

1. Fill in the fit function in 'model.py' with the K-Means algorithm.
2. Fill in the predict function to predict the cluster of each data point.
3. Fill in the replace_with_cluster_centers function which replaces each point with its closest cluster center.
4. Fill in the error function in 'utils.py' to calculate the Mean Squared difference between the original image and the clustered image.
5. Run main.py with different number of clusters ($k = 2, 5, 10, 20, 50$) and save the resulting images and note down the error corresponding to each k.
6. Plot the MSE with respect to k.

Deliverables

- Code for implementing K-Means algorithm without using any libraries except numpy.
- A report containing the details of the algorithm, the resultant images where the pixels are replaced by the nearest cluster centers, and the plot of Mean Squared Error as a function of the number of clusters. (The images and plot should be inside the PDF report as well)

Submission: Email a **single zip file** of the format - `Asst2_FirstName_Last5DigitsOfSRNo.zip` to `e0.270.iisc.2023@gmail.com` with Subject: `Asst2_FirstName_Last5DigitsOfSRNo` on or before the due date. The zip file should contain the following files: main.py, utils.py, model.py, report.pdf, the original and clustered images, and the plot of error vs k.