ELE888 Intelligent Systems

Lab 4: Unsupervised Learning

# Introduction

In this lab, the K-means algorithm for clustering unlabeled data was implemented. The K-means is a type of unsupervised artificial learning where natural clusters within unlabeled data samples can be identified through an iterative learning process. The goal of K-means clustering algorithm is to identify k mean vectors or cluster centres within the specified unlabeled data.

This algorithm will be implemented in the application for finding the dominant colours of an image.

# Theory

## K-means Clustering Algorithm

1. begin initialize *n, c = k,*

2. do classify *n* samples according to nearest

3. recompute

4. until no change in

5. return

6. end

*B. Xie-Beni (XB)*

XB is used to assess the quality the clustering method. The formula is given below:

# Results

*Part (a)*

K-means algorithm when c = 2.

Initial Mean Values:

M0 =

167.3447 239.5471 61.5625

244.9756 117.2189 195.5579

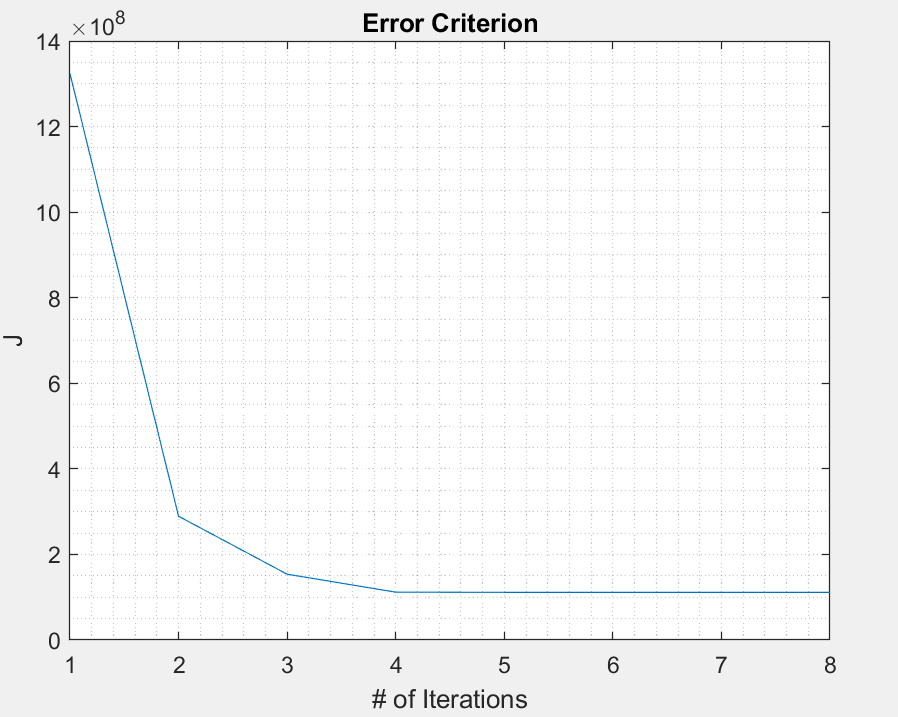


Figure 1. Error Criterion J

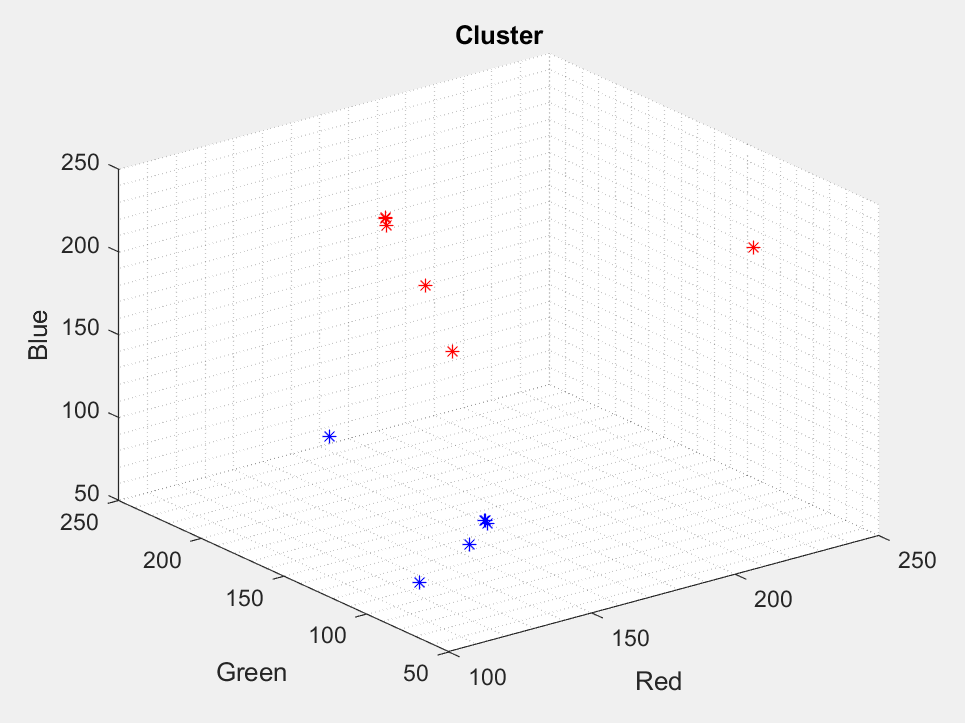


Figure 2. RBG cluster mean

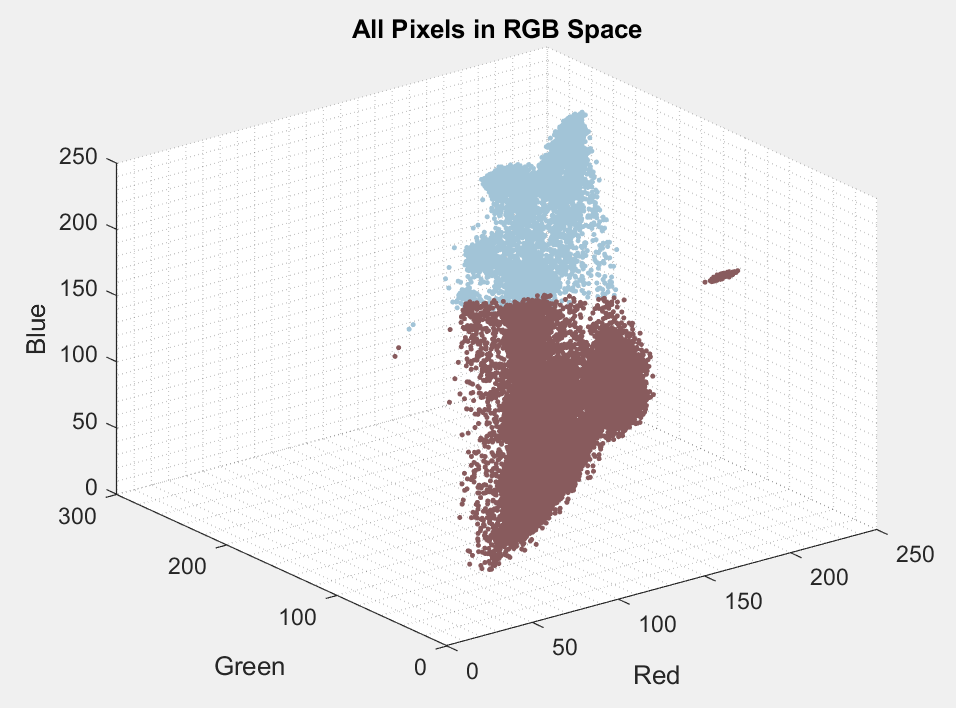


Figure 3. Data Sample in RGB space for c = 2

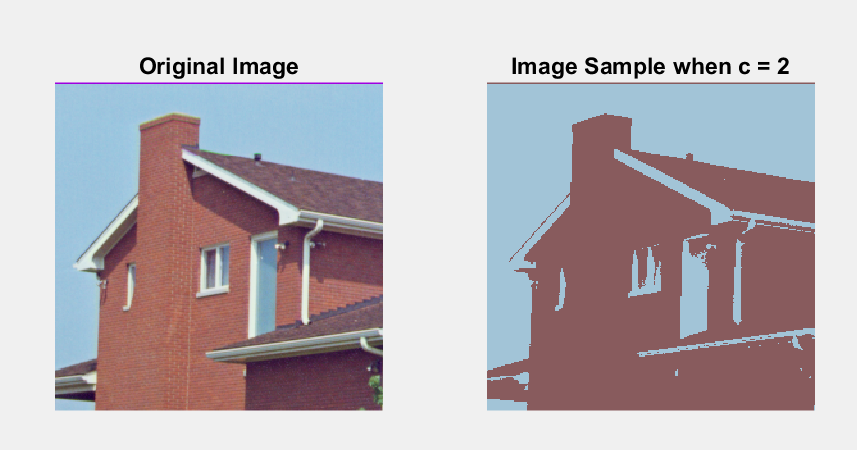


Figure 4. Original image vs sampled image when c = 2

*Part (b)*

K-means algorithm when c = 5.

Initial Mean Values:

M1 =

166.4927 106.4954 96.5852

93.4361 58.0930 73.6996

162.6217 197.6543 217.4457

159.7422 113.5673 222.8516

122.9931 108.1533 114.4409

M2 =

245.9278 167.1520 203.4467

110.2837 27.9875 124.3390

177.1618 238.1088 196.0844

193.3153 47.8025 100.9817

110.3238 67.8756 69.5994



Figure 5. Original image vs sampled image when c = 5

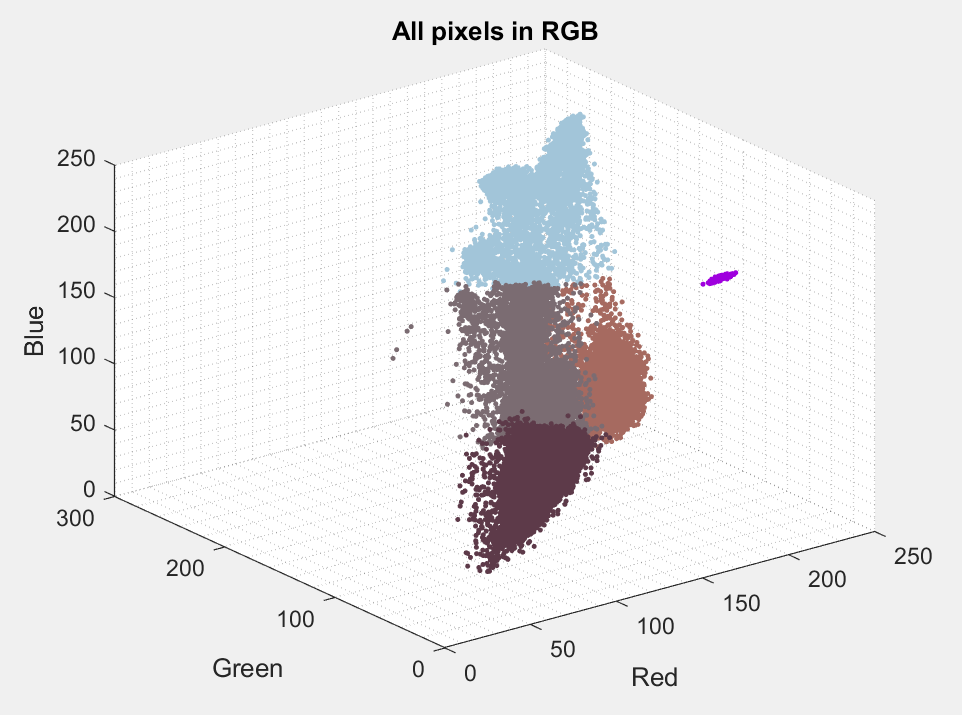


Figure 6. Data Sample in RGB space for c = 2 (First run with M1)

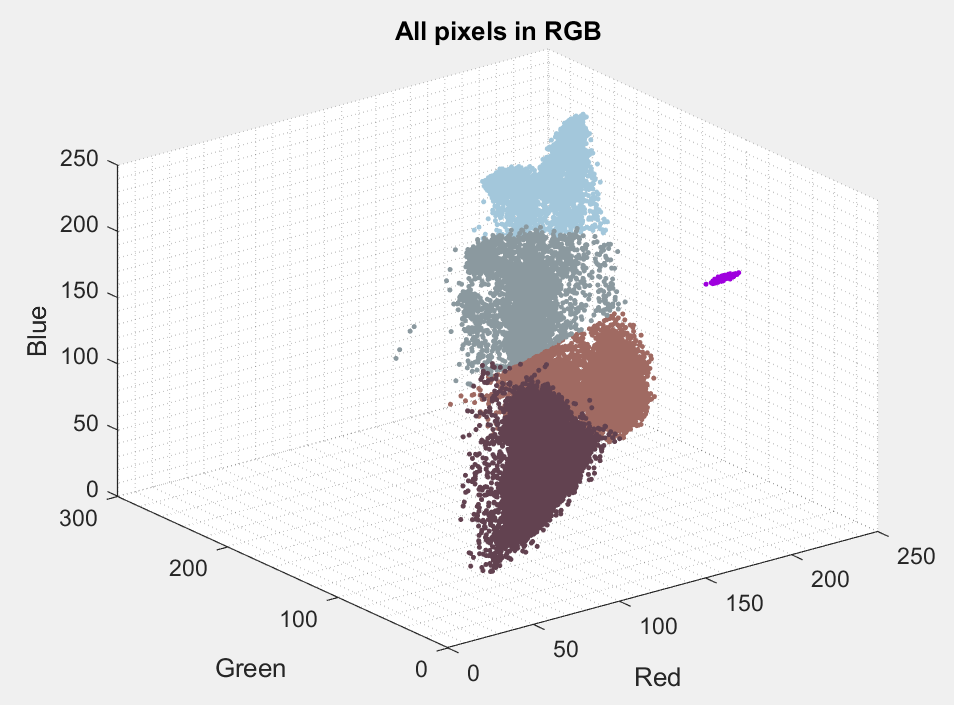


Figure 7. Data Sample in RGB space for c = 2 (First run with M2)

*Part (c)*

Calculate the XB to assess quality of clustering

XB1 = 0.2560

XB2 = 0.2381

# Conclusion/Discussion

It is clearly seen by comparing Figure 4 and Figure 5 that with larger number of k, we are able to pull out more dominant colours from the image and construct with more colour accuracy with the original image. This will result in a smaller XB value meaning better performance.. Furthermore, from Figure 7 and Figure 8 that with different initial means, can result in different clustered regions. This can affect the quality of the clustering as seen from XB1 and XB2. The XB for the 2 different are slightly different.

In conclusion, the K-means clustering algorithm is a great tool for cluster analysis. It can be very accurate in classifying unlabeled data.

Reference

[1] N. Zhang, ”ELE888/EE8209 – Intelligent Systems – Student Lab Manual,” Department of Electrical and Computer Engineering, Ryerson University, Toronto, Ontario, March 2019.