

|  |  |
| --- | --- |
| **Course Title:** | **Intelligent Systems** |
| **Course Number:** | **ELE 888** |
| **Semester/Year (e.g.F2016)** | **W2021** |
|  |  |
| **Instructor:** | **Dr. Xiao-Ping Zhang** |

|  |  |
| --- | --- |
| *Assignment/Lab Number:* | **Lab 4** |
| *Assignment/Lab Title:* | **Unsupervised Learning** |

|  |  |
| --- | --- |
| *Submission Date:* | **April-11-2021** |
| *Due Date:* | **April-11-2021** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student**  **LAST Name** | **Student**  **FIRST Name** | **Student Number** | **Section** | **Signature\*** |
| **An** | **Nicholas** | **500750378** | **3** | **NA** |
| **Parmar** | **Gurvir** | **500765786** | **1** | **GP** |
| **Shreekant** | **Vatsal** | **500771363** | **1** | **VS** |

*\*By signing above you attest that you have contributed to this written lab report and confirm that all work you have contributed to this lab report is your own work. Any suspicion of copying or plagiarism in this work will result in an investigation of Academic Misconduct and may result in a “0” on the work, an “F” in the course, or possibly more severe penalties, as well as a Disciplinary Notice on your academic record under the Student Code of Academic Conduct, which can be found online at:* [*http://www.ryerson.ca/senate/current/pol60.pdf*](http://www.ryerson.ca/senate/current/pol60.pdf)

**Objective:**

In unsupervised learning, natural clusters within unlabeled data samples (i.e. with no categorical information) may be identiﬁed using an iterative learning process. When the functional form of the underlying probability densities of the data are assumed to be known, the only thing that must be learnt is the value of an unknown parameter vector.

One elementary but popular approximate method that performs the above is the k-means clustering algorithm. The goal of the k-means clustering algorithm is to identify k mean vectors or cluster centres within the given unlabeled data.

**Observation:**

**Part A:**

Chart, line chart

Description automatically generated

Fig 1: Error Criterion J Plot

Chart, radar chart

Description automatically generated

Fig 2: Cluster Means Plot for The Two Stages of The Clustering Process

Table 1: Initial RGB Mean Values with c = 2

|  |  |  |
| --- | --- | --- |
| Red | Green | Blue |
| 162.1561 | 196.6642 | 216.1052 |
| 136.3930 | 91.4700 | 93.6969 |

Table 2: Final RGB Mean Values with c = 2

|  |  |  |
| --- | --- | --- |
| Red | Green | Blue |
| 195.2068 | 47.6525 | 113.6245 |
| 202.7760 | 124.8899 | 164.8098 |

Chart, radar chart

Description automatically generated

Fig 3: Labeled Data Samples Plot (pixels) in RGB Space

Logo, company name

Description automatically generated

Fig 4: Original Image in Mean Colors

**Part B:**

**Run 1:**

Chart, line chart

Description automatically generated

Fig 5: Error Criterion J Plot

Chart, radar chart

Description automatically generated

Fig 6: Cluster Means Plot for The Two Stages of The Clustering Process

Table 3: Initial RGB Mean Values with c = 5

|  |  |  |
| --- | --- | --- |
| Red | Green | Blue |
| 178.2646 | 38.0700 | 62.0989 |
| 227.1803 | 65.6646 | 236.9622 |
| 244.6193 | 214.3829 | 89.2459 |
| 139.5400 | 64.8420 | 50.1318 |
| 35.3492 | 207.6426 | 64.0264 |

Table 4: Final RGB Mean Values with c = 5

|  |  |  |
| --- | --- | --- |
| Red | Green | Blue |
| 159.7422 | 0 | 222.8516 |
| 96.1283 | 62.8571 | 77.4643 |
| 156.9644 | 194.5660 | 216.9730 |
| 212.8231 | 221.9035 | 215.9856 |
| 157.0964 | 107.8332 | 101.4552 |

Chart, radar chart

Description automatically generated

Fig 7: Labeled Data Samples Plot (pixels) in RGB Space

**Run 2:**

Chart, line chart

Description automatically generated

Fig 8: Error Criterion J Plot

Chart, radar chart

Description automatically generated

Fig 9: Cluster Means Plot for The Two Stages of The Clustering Process

Table 5: Initial RGB Mean Values with c = 5

|  |  |  |
| --- | --- | --- |
| Red | Green | Blue |
| 140.2144 | 130.8786 | 46.8965 |
| 36.9635 | 102.4610 | 61.1879 |
| 217.5229 | 19.3715 | 106.4031 |
| 158.6241 | 61.1786 | 12.6619 |
| 89.4929 | 31.4463 | 230.1926 |

Table 6: Final RGB Mean Values with c = 5

|  |  |  |
| --- | --- | --- |
| Red | Green | Blue |
| 166.4733 | 106.6846 | 96.7292 |
| 93.0389 | 57.5181 | 73.2731 |
| 123.1519 | 103.1813 | 116.1449 |
| 157.2345 | 194.8986 | 217.3104 |
| 212.8629 | 221.9088 | 215.9715 |

Diagram, schematic, radar chart

Description automatically generated

Fig 10: Labeled Data Samples Plot (pixels) in RGB Space

**Part C:**

The Xie-Beni index for c=2 and c=5(both runs) was as follows:

C = 2: XB(2) = 0.5352

C = 5: XB(5) = 0.0679 (First Run)

C = 5: XB(5) = 0.0578 (Second Run)

The Xie-Beni index will always be changing because the initial mean values for i are constantly adjusting. During the run 2 when c was equal to 5, it was observed that the XB values were close and very comparable. Implementing more clusters of mean values would lead to smaller XB values and thus better performance of the K-means algorithm.

**Conclusion**

The implemented K-means algorithm is a great tool for calculating the distances to assign a point to a cluster. It is very precise in minimizing the sum of distances between the points and their respective cluster centroid. This was observed by comparing the original image to the image in cluster means color, the Xie-Beni index or the labelled cluster mean plots in various iterations.

The lack of optimization was an issue in terms of the computation time of the algorithm. If a sample input were fed to the model with a large pixel size, the processing power would take a hit. Although even without the optimization parameters, the model takes a couple of minutes to run.