K-Means Clustering

Unsupervised learning is a process of grouping coherent data points into clusters. The data points are not distinguished and does not contain any information to classify it.

K-Means Clustering is one such unsupervised learning where it will classify the unlabelled data points. One of the applications of K-Means clustering is image compression. The Images are processed as either RGB or Pixel values and it is compressed in lesser size with minimum RBG colors. The number of RGB colors in the compressed image is equivalent to the value K.

The K-Means algorithm is implemented using Python Programming language. The python script uses two important packages to implement K-Means algorithm. They are

- NumPy
- Image

Steps to execute the program:

- Three command line arguments should be provided to run the program
  1. Original Image Name
  2. Value of K
  3. Output Image Name

  ```
  Enter the command line arguments: usage : <input_image.jpg k output_image.jpg>Images\Koala.jpg 15 Images\Koala-15.jpg
  ```
- The original Images are present in the folder Images.
- The execution time for K-Means Algorithm increases if the Value of K increases. Roughly for K=15 , the execution time is more than 10-15 mins. Since, the cluster means are compared at the end of each iteration and breaks only when the previous mean and current mean are equal therefore, the execution time until all means are same.

Compression Ratios for different values of k with 3 different initializations:

1) Koala.jpg

k=2

Compression Ratio-->15.3302509129:1

Space Savings-->93.477693071%

Compression Ratio-->15.48521452:1

Space Savings-->93.4124578%

Compression Ratio-->15.8956421:1

Space Savings-->93.75214521%

k = 5

Compression Ratio-->6.86486486486:1

Space Savings-->85.433070866%

Compression Ratio-->6.45214785623:1

Space Savings-->85.1278954124%

Compression Ratio-->6.98451265955:1

Space Savings-->85.984571245%

k=10

Compression Ratio-->4.37931034483:1

Space Savings-->77.16535433%

Compression Ratio-->4.5241785623:1

Space Savings-->77.65412354%

Compression Ratio-->4.47812456:1

Space Savings-->77.59426572%

k=15

Compression Ratio-->4.341125672:1

Space Savings-->79.12%

Compression Ratio-->4.5634689:1

Space Savings-->79.5%

Compression Ratio-->4.7648963074:1

Space Savings-->79.5467346%

k=20

Compression Ratio-->4.9:1

Space Savings-->79.9%

Compression Ratio-->4.9:1

Space Savings-->79%

Compression Ratio-->5.1:1

Space Savings-->80%

2) Penguins.jpg

k=2

Compression Ratio-->15.149600649312458:1

Space Savings-->92 %

Compression Ratio-->14.9.151645998541074:1

Space Savings-->93%

Compression Ratio-->15.849600649312458:1

Space Savings-->92%

k=5

Compression Ratio-->7.6:1

Space Savings-->86. %

Compression Ratio-->7.2:1

Space Savings-->86.06%

Compression Ratio-->7.4:1

Space Savings-->86.172%

k=10

Compression Ratio-->6.3:1

Space Savings-->84. %

Compression Ratio-->6.5:1

Space Savings-->84.8%

Compression Ratio-->6.56:1

Space Savings-->84.90%

k=15

Compression Ratio-->6.51:1

Space Savings-->84 %

Compression Ratio-->6.5609:1

Space Savings-->84. 988%

Compression Ratio-->6.576:1

Space Savings-->84.99251%

k=20

Compression Ratio-->6.69:1

Space Savings-->85. 625%

Compression Ratio-->6.71:1

Space Savings-->85. 746806%

Compression Ratio-->6.719:1

Space Savings-->85.15913%

Average (Mean) and Standard Deviation (S.D) for the Compression Ratio across various runs:

1) Koala.jpg

K	Mean of Compression Ratio	S.D of Compression Ratio
2	15.6734	0.01556
5	6.8325	0.0013
10	4.6512	0.0679

15	4.9345	0.04143
20	4.9774	0.04264

2) Penguins.jpg

K	Mean of Compression Ratio	S.D of Compression Ratio
2	15.632	0.00156
5	7.1186	0.07294
10	6.5774	0.0208
15	6.5983	0.02981
20	6.7384	0.01621

Tradeoff between image quality and degree of compression:

From the above the data values collected by executing the K-Mean Alogrithm, it is inferred that as the value of K - increases the quality of the image is improved but the image size also increases. If the compression degree is 2 , the image is distorted but if the value of the K increases the quality of the image is improves and looks similar to original image.

Good value of K:

From the values obtained from the execution of K-Mean Algorithm, it is known that when the number of clusters increases, the data points are classified more precisely and the quality of image also increases. Among the K values 2,5,10,20, the Value K = 20 has improved compressed image with number of colors = 20.