

Name: Naveenraj Palanisamy

NetId: NXP154130

Machine Learning Assignment (CS 6375.001)-(K Means)

(Note: Runs in python 3.5)

(Note: File Path should be full path)

Folder Structure:

This PC > OS (C:) > 3 > Assignment3 > K Means				
<input type="checkbox"/>	Name	Date modified	Type	Size
	KMeans.java	3/27/2016 8:49 PM	JAVA File	2 KB
	Koala.jpg	3/27/2016 8:49 PM	JPG File	763 KB
	npx154130_kmeans.docx	4/16/2016 9:54 PM	Microsoft Word D...	12 KB
	Penguins.jpg	3/27/2016 8:49 PM	JPG File	760 KB

KMeans is the java file that we need to run.

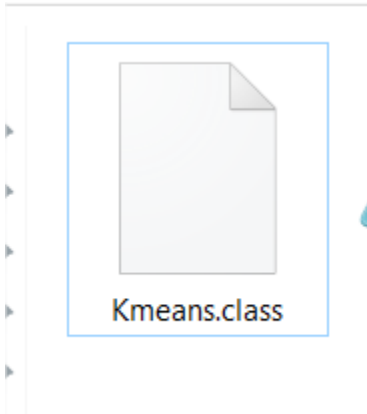
2 Images files are the input image files.

Steps to run:

- 1) Open command prompt
- 2) Go to the working directory in command prompt
- 3) Set path and class path
- 4) Set path to java home and class path to your working directory.

```
C:\3\Assignment3\K_Means>set path=C:\Program Files\Java\jdk1.8.0_77\bin;
C:\3\Assignment3\K_Means>set classpath=.;
C:\3\Assignment3\K_Means>
```

- 5) Type `javac Kmeans.Java`
- 6) This will generate the `.class` file, we can run this `.class` file using `java` command.



7) Run this by Java Kmeans Koala.jpg 2 Koala_2_cluster.jpg

Run:

- 1) Javac Kmeans.java
- 2) Java Kmeans Penguins.jpg 2 Penguins_2_clusters.jpg

```
C:\3\Assignment3\K_Means>Javac Kmeans.java
Note: Kmeans.java uses unchecked or unsafe operations.
Note: Recompile with -Xlint:unchecked for details.
```

Images after Compression:

Image: Penguins.jpg

Original Image:



Clusters: 2



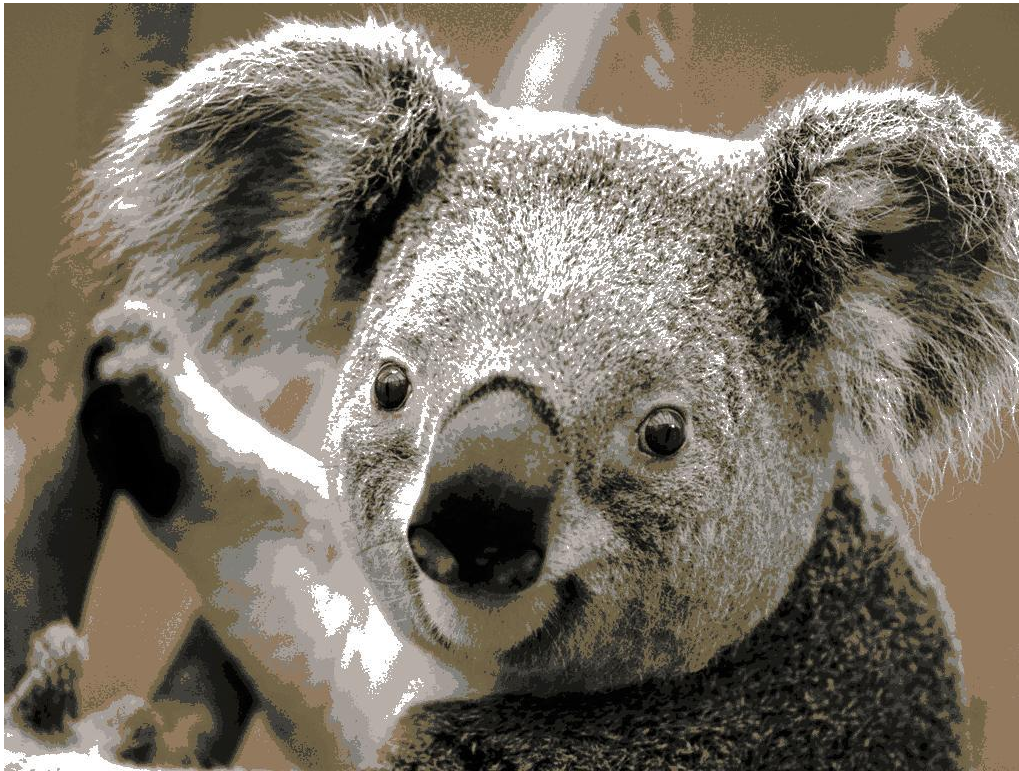
Cluster: 5



Cluster 10:



Cluster 15:



Cluster 20:

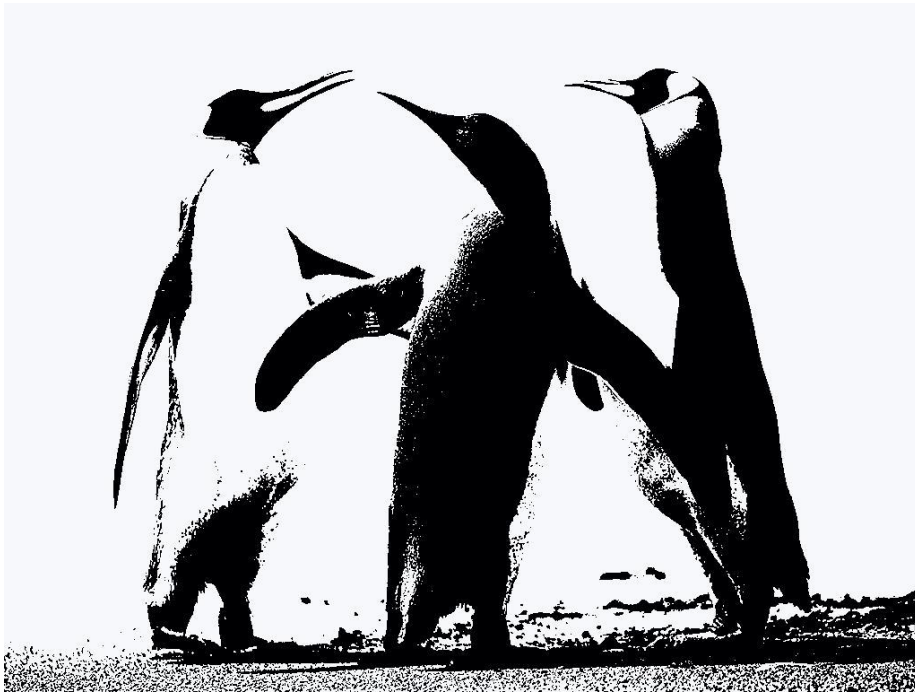


Image: Penguins.jpg

Original Image:



Cluster 2:



Cluster 5:



Cluster 10:



Cluster 15:



Cluster 20:



Comparison ratios for different K values:

Image: Koala.jpg

1)

K value: 2

Initialization values:

```
Cluster 0 R value 28 G value 242 B value 223  
Cluster 1 R value 78 G value 97 B value 206
```

Comparison ratio: 762kb/191kb: 3.98

2)

K value: 2

Initialization values:

```
C:\5\Assignments\K_Means>java KMeans koala.jpg
Cluster 0 R value 251 G value 76 B value 150
Cluster 1 R value 68 G value 202 B value 108
```

Comparison ratio: 762kb/190kb: 3.97

3)

K value: 10

Initialization values:

```
C:\5\Assignments\K_Means>java KMeans koala.jpg
Cluster 0 R value 26 G value 152 B value 188
Cluster 1 R value 116 G value 156 B value 106
Cluster 2 R value 193 G value 27 B value 221
Cluster 3 R value 159 G value 154 B value 35
Cluster 4 R value 38 G value 109 B value 142
Cluster 5 R value 166 G value 146 B value 102
Cluster 6 R value 168 G value 200 B value 21
Cluster 7 R value 250 G value 53 B value 45
Cluster 8 R value 126 G value 238 B value 110
Cluster 9 R value 57 G value 103 B value 98
```

Comparison ratio: 762kb/175kb: 3.354

4)

K value: 20

Initialization values:

```
C:\5\Assignments\K_Means\java Kmeans Koala.jpg
Cluster 0 R value 111 G value 224 B value 101
Cluster 1 R value 169 G value 34 B value 148
Cluster 2 R value 153 G value 173 B value 49
Cluster 3 R value 249 G value 70 B value 137
Cluster 4 R value 211 G value 46 B value 194
Cluster 5 R value 69 G value 224 B value 238
Cluster 6 R value 107 G value 189 B value 84
Cluster 7 R value 176 G value 30 B value 147
Cluster 8 R value 202 G value 197 B value 1
Cluster 9 R value 32 G value 33 B value 254
Cluster 10 R value 64 G value 98 B value 53
Cluster 11 R value 190 G value 24 B value 25
Cluster 12 R value 178 G value 242 B value 85
Cluster 13 R value 214 G value 215 B value 143
Cluster 14 R value 14 G value 204 B value 186
Cluster 15 R value 49 G value 189 B value 13
Cluster 16 R value 246 G value 108 B value 229
Cluster 17 R value 26 G value 237 B value 55
Cluster 18 R value 131 G value 159 B value 14
Cluster 19 R value 148 G value 83 B value 215
```

Comparison ratio: 762kb/162kb: 4.7037

Similarly done for the for different K values several times.

Image: Penguins.jpg

1)

K value: 2

Initialization values:

```
Cluster 0 R value 193 G value 77 B value 192  
Cluster 1 R value 61 G value 117 B value 155
```

Comparison ratio: 759kb/98.8kb: 7.712

2)

K value: 10

Initialization values:

```
C:\3\Assignment3\K_Means>java Kmeans Penguins  
Cluster 0 R value 243 G value 192 B value 222  
Cluster 1 R value 3 G value 131 B value 43  
Cluster 2 R value 156 G value 113 B value 131  
Cluster 3 R value 90 G value 144 B value 8  
Cluster 4 R value 230 G value 143 B value 107  
Cluster 5 R value 122 G value 222 B value 124  
Cluster 6 R value 104 G value 58 B value 74  
Cluster 7 R value 247 G value 141 B value 0  
Cluster 8 R value 204 G value 104 B value 109  
Cluster 9 R value 9 G value 43 B value 48
```

Comparison ratio: 759kb/119kb: 6.37

3)

K value: 20

Initialization values:

```
Cluster 0 R value 88 G value 225 B value 97
Cluster 1 R value 62 G value 189 B value 181
Cluster 2 R value 220 G value 8 B value 96
Cluster 3 R value 244 G value 83 B value 18
Cluster 4 R value 130 G value 40 B value 118
Cluster 5 R value 247 G value 248 B value 119
Cluster 6 R value 141 G value 244 B value 192
Cluster 7 R value 37 G value 88 B value 124
Cluster 8 R value 242 G value 54 B value 74
Cluster 9 R value 183 G value 141 B value 221
Cluster 10 R value 72 G value 214 B value 126
Cluster 11 R value 42 G value 176 B value 22
Cluster 12 R value 241 G value 193 B value 66
Cluster 13 R value 155 G value 214 B value 163
Cluster 14 R value 211 G value 96 B value 206
Cluster 15 R value 190 G value 156 B value 204
Cluster 16 R value 134 G value 252 B value 61
Cluster 17 R value 122 G value 25 B value 83
Cluster 18 R value 226 G value 255 B value 86
Cluster 19 R value 61 G value 242 B value 128
```

Comparison ratio: 759kb/111kb: 6.837

Similarly done for the for different K values several times.

Image 1: Koala.jpg

Average Compression rate: 3.8

Compression rate varies by 0.3 for different K values.

Image 1: Penguins.jpg

Average Compression rate: 6.8

Compression rate varies by 0.46 for different K values.

Is there a tradeoff between image quality and degree of compression? What would be a good value of K for each of the two images?

Actually it depends upon the initial assignment, if the initial assignment is proper we can get better image quality with high degree of compression.

For the supplied images:

Image: Koala.jpg

Good K value seems to be 20, it produces better image with higher degree of compression.

Image: Penguins.jpg

Good K values seems to be 50, it produces better image with higher degree of compression.

Note: (Attached this image for reference)

Notes:

- 1) Increase of K values will increase number of colors in the image, Tested this with 1000 values for K and got the exact image for the image that had 1000 different colors.