CS 6375: MACHINE LEARNING

Homework 5.1

K-MEANS CLUSTERING and EM ALGORITHM

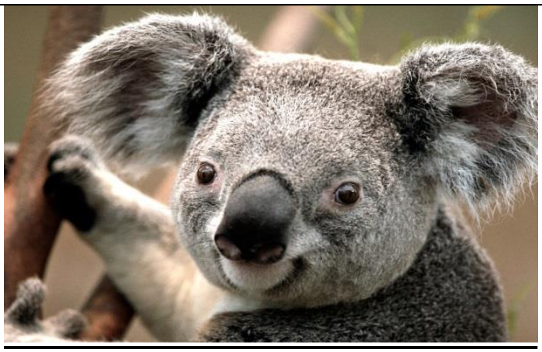
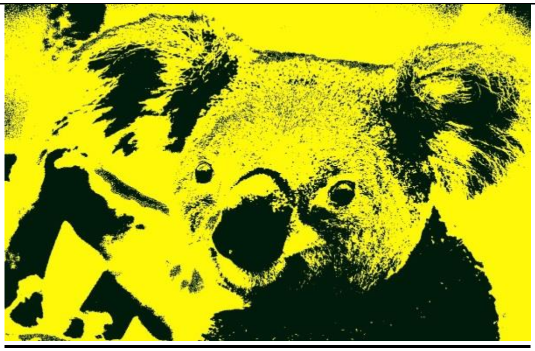
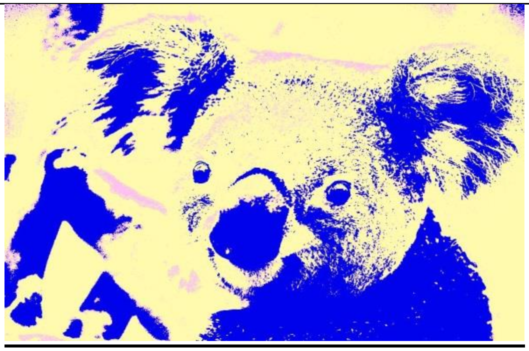
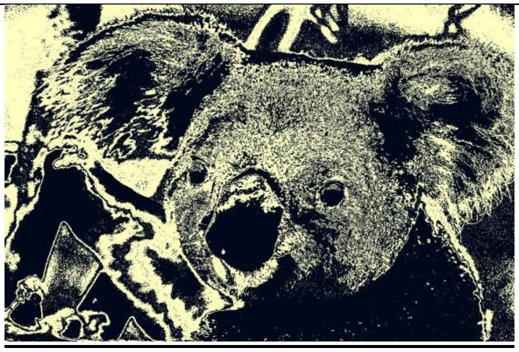
Date: 12/04/2013

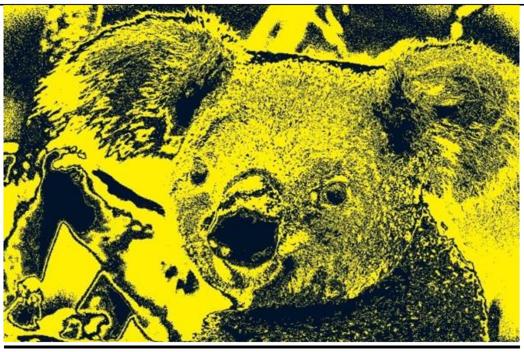
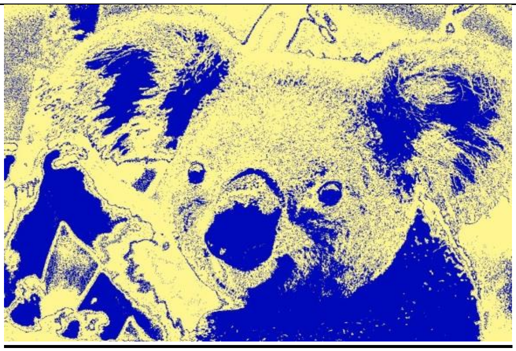
-Sudarshan Sudarshan

-sxs121433

K-MEANS CLUSTERING ON IMAGES:

Image # 1: Koala

Sl. No.	Image	K	Compression Ratio
1.		Original	
2.		K = 2	**3.40 : 1** **(**CR = 762 / 224**)**
3.		K=5	**3.98 : 1** **(**CR = 762 / 191**)**
4.		K=10	**2.06 : 1** **(**CR = 762 / 369**)**

5.		**K = 15**	**1.79 : 1** **(**CR = 762 / 425**)**
6.		**K= 20**	**2.15 : 1** **(**CR = 762 / 354**)**

Compression Ratio Calculation Formula:

$$\textbf{Compression Ratio} = \frac{\textbf{Uncompressed Size}}{\textbf{Compressed Size}}$$

Ratio is represented as: **Uncompressed: compressed (Ex: 5:1)**

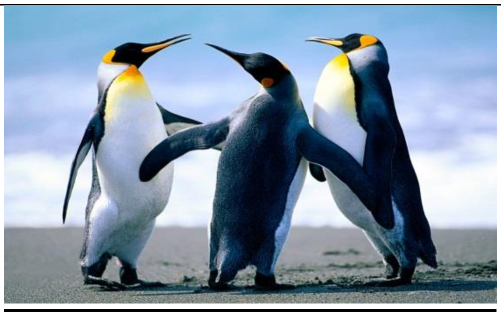
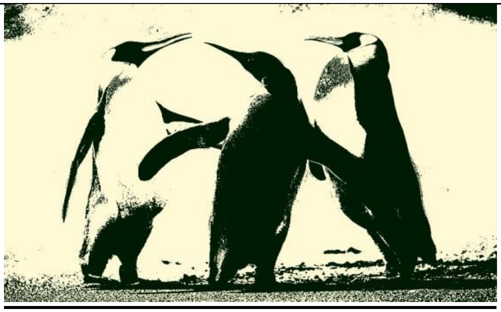
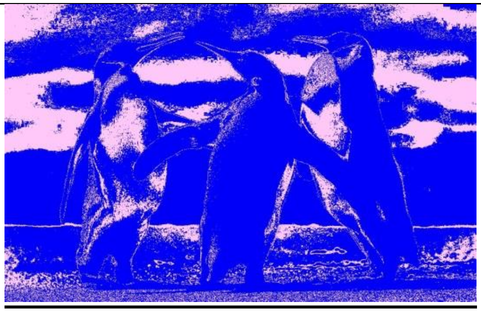
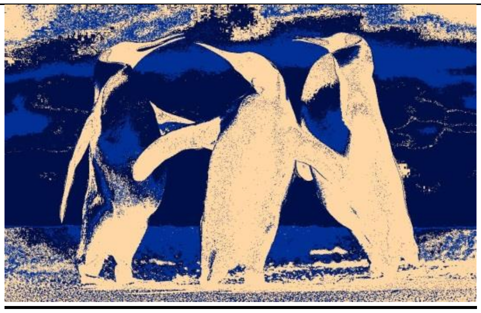
Results obtained after 3 repetitions:

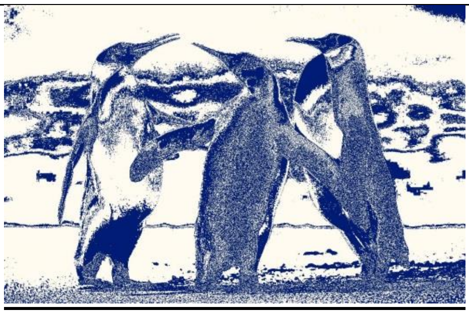
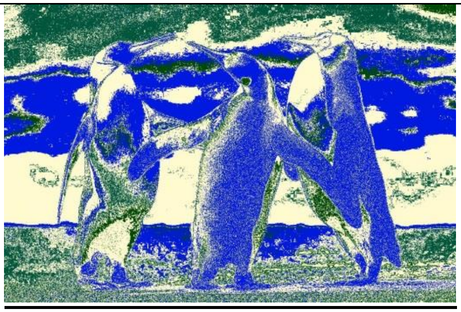
	Compression Ratio			
	Repetition 1	**Repetition 2**	**Repetition 3**	**Average**
K = 2	3.25:1	3.17:1	3.37:1	3.26:1
K = 5	3.98:1	2.49:1	2.00:1	2.82:1
K = 10	2.06:1	2.20:1	1.87:1	2.04:1
K = 15	1.79:1	1.55:1	1.83:1	1.72:1
K = 20	2.15:1	1.95:1	1.56:1	1.88:1

Conclusion:

There is a trade-off between the image quality and the degree of compression. For this image K=10, has given a good image when compared to others. There can be a better result, as K-Means only converges to local minima and also, the outcome depends on the K-Initializations.

Image # 2: Penguins

Sl. No.	Image	K	Compression Ratio
1.		Original	
2.		K = 2	**6.54 : 1** **(**CR = 759 / 116**)**
3.		K=5	**3.06 : 1** **(**CR = 759 / 248**)**
4.		K=10	**3.08 : 1** **(**CR = 759 / 246**)**

5.		**K = 15**	**2.46 : 1** **(**CR = 759 / 308**)**
6.		**K= 20**	**1.89 : 1** **(**CR = 759 / 400**)**

Compression Ratio Calculation Formula:

$$\textbf{Compression Ratio} = \frac{\textbf{Uncompressed Size}}{\textbf{Compressed Size}}$$

Ratio is represented as: **Uncompressed: compressed (Ex: 5:1)**

Results obtained after 3 repetitions:

	Compression Ratio			
	Repetition 1	**Repetition 2**	**Repetition 3**	**Average**
K = 2	6.54:1	6.12:1	5.42:1	6.02:1
K = 5	3.06:1	4.46:1	4.00:1	3.84:1
K = 10	3.08:1	2.63:1	3.17:1	2.96:1
K = 15	2.46:1	2.01:1	1.83:1	2.10:1
K = 20	1.89:1	1.90:1	1.93:1	1.90:1

Conclusion:

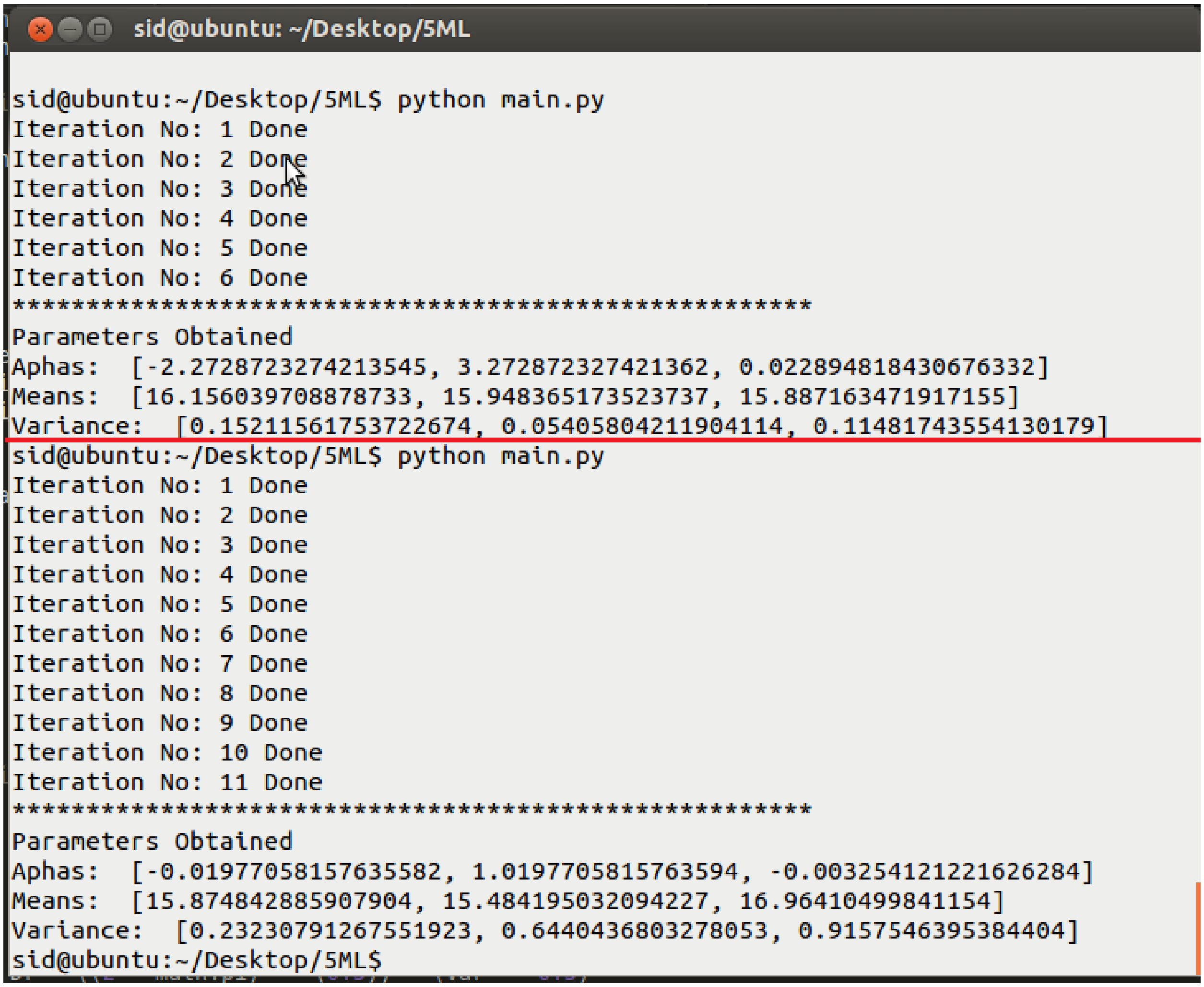
There is a trade-off between the image quality and the degree of compression. For this image K=2, has given a good image when compared to others. There can be a better result, as K-Means only converges to local minima and also, the outcome depends on the K-Initializations.

Note: I have included the K-Initialization taken into consideration in two separate files in the zip folder, viz. *Koala_k_Values.txt* for Koala image and *Penguins_K_Values.txt* for Penguins Image.

EM ALGORITHM:

Initialization taken into consideration: Selected the initial K Gaussian means by randomly selecting K initial data points, and selected the initial K covariance as all being some multiple of the overall data covariance

Experiment 1: For General GMM

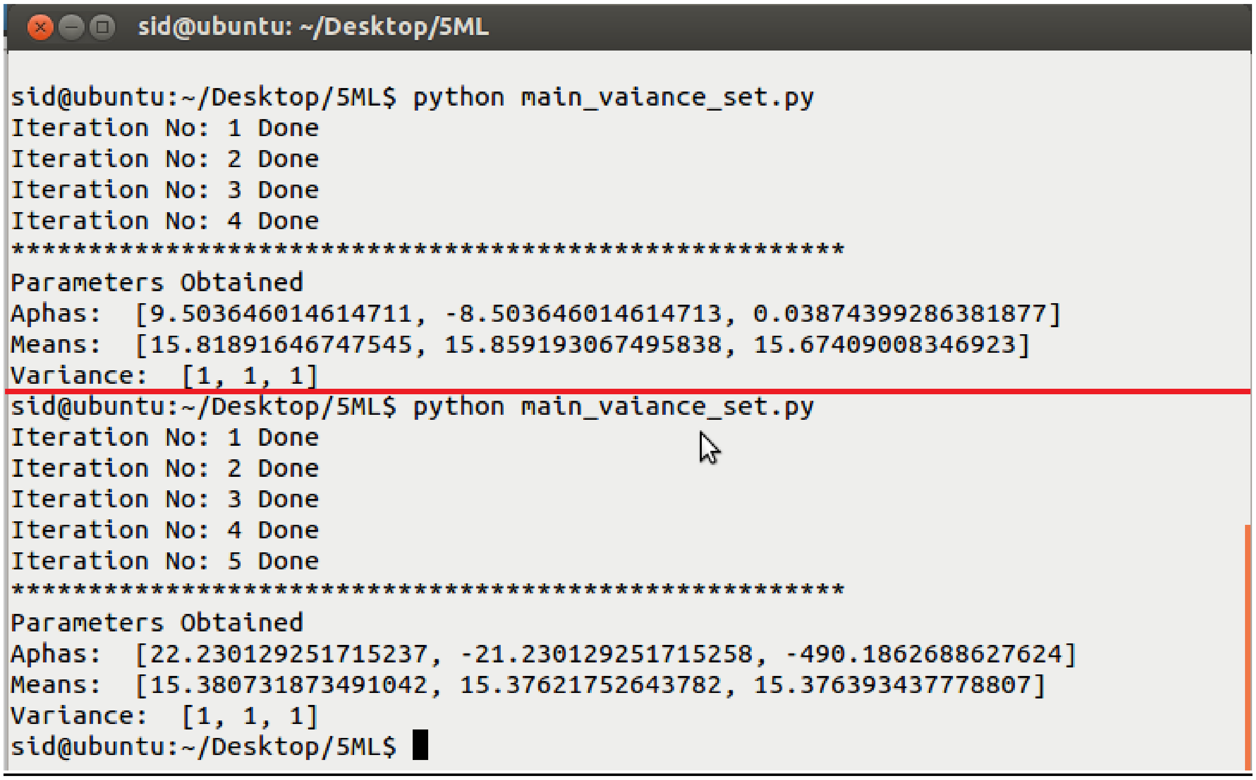


```
sid@ubuntu:~/Desktop/5ML$ python main.py
Iteration No: 1 Done
Iteration No: 2 Done
Iteration No: 3 Done
Iteration No: 4 Done
Iteration No: 5 Done
Iteration No: 6 Done
*******************************************************
Parameters Obtained
Aphas:  [-2.2728723274213545, 3.272872327421362, 0.022894818430676332]
Means:  [16.156039708878733, 15.948365173523737, 15.887163471917155]
Variance:  [0.15211561753722674, 0.05405804211904114, 0.11481743554130179]
sid@ubuntu:~/Desktop/5ML$ python main.py
Iteration No: 1 Done
Iteration No: 2 Done
Iteration No: 3 Done
Iteration No: 4 Done
Iteration No: 5 Done
Iteration No: 6 Done
Iteration No: 7 Done
Iteration No: 8 Done
Iteration No: 9 Done
Iteration No: 10 Done
Iteration No: 11 Done
*******************************************************
Parameters Obtained
Aphas:  [-0.01977058157635582, 1.0197705815763594, -0.003254121221626284]
Means:  [15.874842885907904, 15.484195032094227, 16.96410499841154]
Variance:  [0.23230791267551923, 0.6440436803278053, 0.9157546395384404]
sid@ubuntu:~/Desktop/5ML$
```

Figure 1: Results obtained for Experiment 1

The results/ parameters obtained for Experiment 1 is shown in the Figure 1. The figure shows two different runs (separated by red line). For each run, No. of iterations, the alpha parameters, the means and the variance are also depicted.

Experiment 2: For General GMM with Variance set to 1



```
sid@ubuntu:~/Desktop/5ML$ python main_vaiance_set.py
Iteration No: 1 Done
Iteration No: 2 Done
Iteration No: 3 Done
Iteration No: 4 Done
******************************************************
Parameters Obtained
Aphas:  [9.503646014614711, -8.503646014614713, 0.03874399286381877]
Means:  [15.81891646747545, 15.859193067495838, 15.67409008346923]
Variance:  [1, 1, 1]
sid@ubuntu:~/Desktop/5ML$ python main_vaiance_set.py
Iteration No: 1 Done
Iteration No: 2 Done
Iteration No: 3 Done
Iteration No: 4 Done
Iteration No: 5 Done
******************************************************
Parameters Obtained
Aphas:  [22.230129251715237, -21.230129251715258, -490.1862688627624]
Means:  [15.380731873491042, 15.37621752643782, 15.376393437778807]
Variance:  [1, 1, 1]
sid@ubuntu:~/Desktop/5ML$
```

Figure 2: Results obtained for Experiment 2

The results/ parameters obtained for Experiment 2 is shown in the Figure 2. The figure shows two different runs (separated by red line). For each run, No. of iterations, the alpha parameters, the means and the variance are also depicted.

Stopping condition Used: Log Likelihood

Conclusion:

As K-means, even EM Algorithm converges to local minima. There might be cases, where the results obtained are better than the observed results listed in this report, having said that, there might also be cases, where the results might be unsatisfactory.