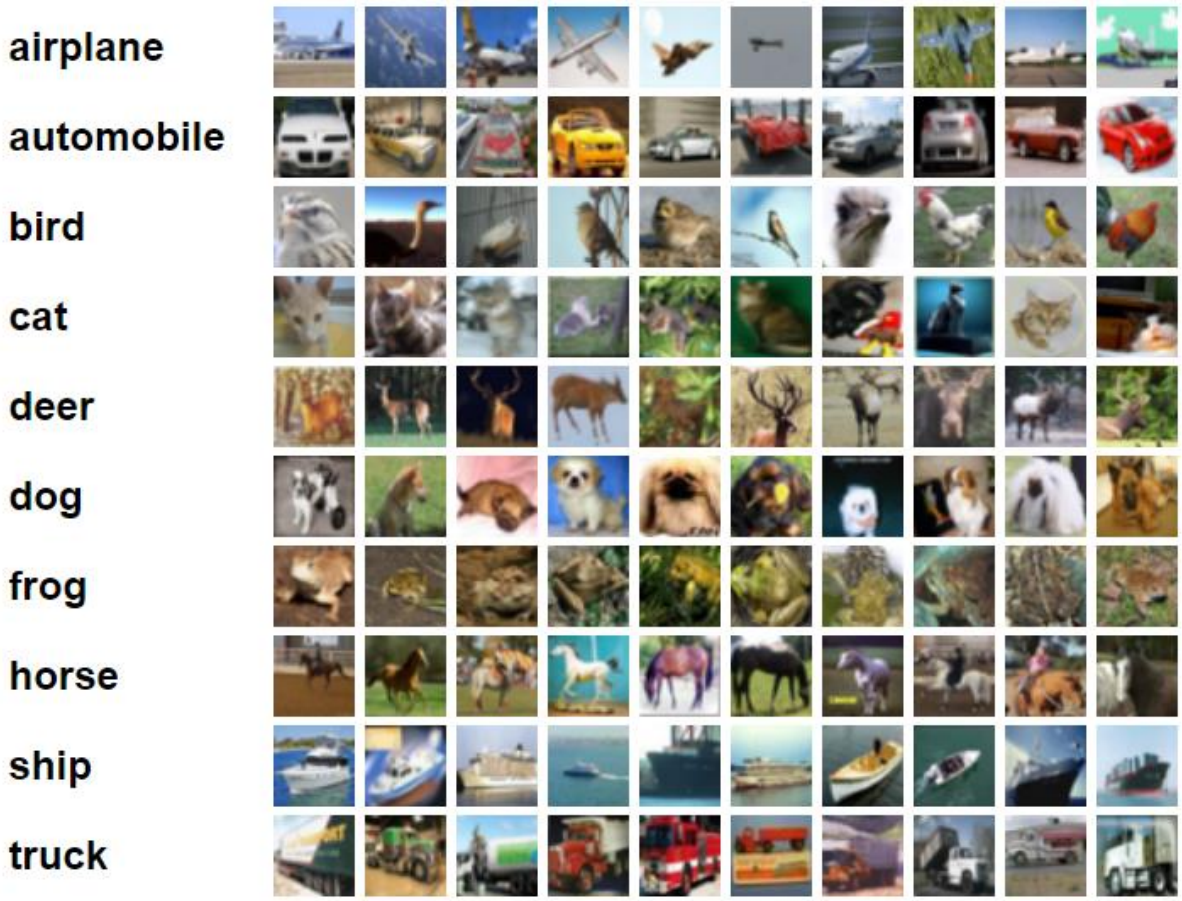
K – MEANS CLUSTERING

Youssef Hussein Abdelsalam

5459

Artificial Intelligence CC482

CIFAR – 10 IMAGE DATASET



Code Output Explanation

A new folder named sample is created for storing the results for clustering. K folders are created inside “sample” named

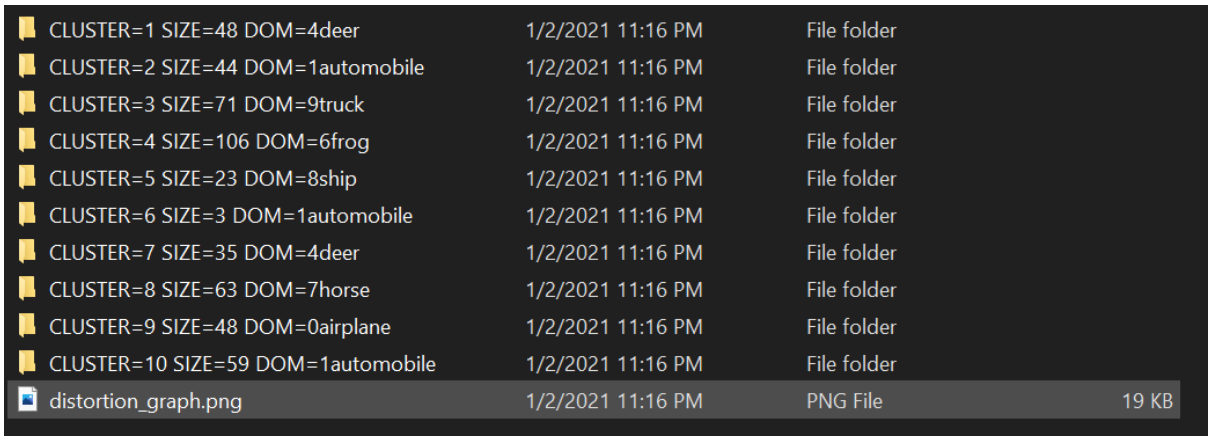
“CLUSTER=i SIZE=size_of_images_in_this_cluster DOM=#_of_dominant_class_and_name_of_class”

Classes:

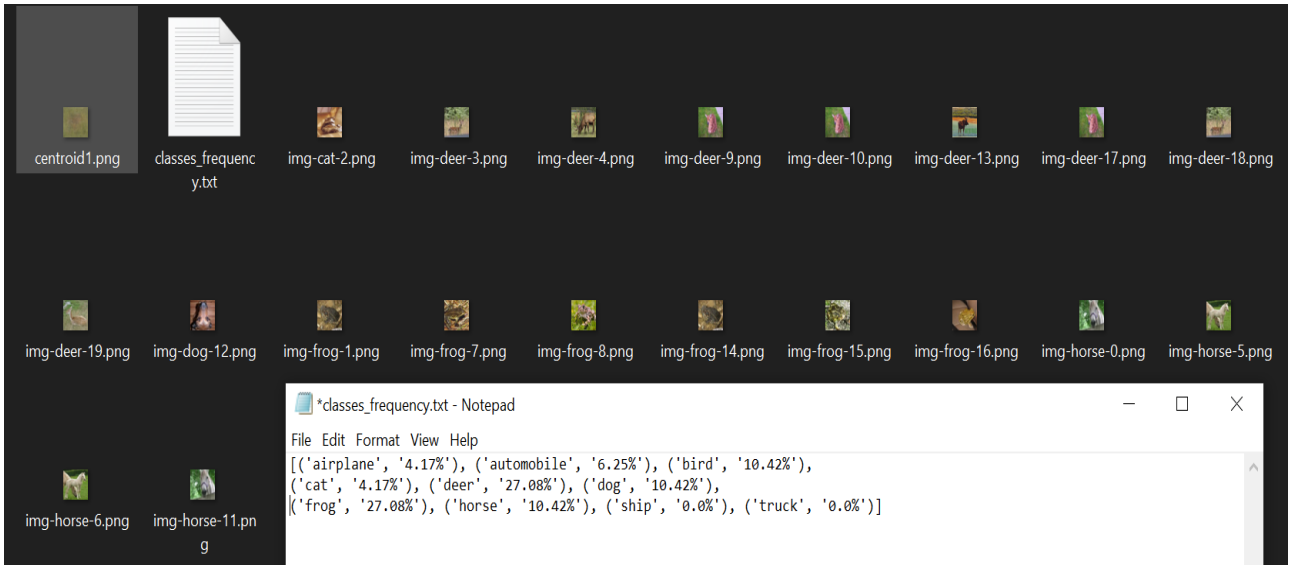
0->airplane, 1->automobile, 2->bird, 3->cat, 4->deer, 5->dog, 6->frog, 7->horse, 8->ship, 9->truck

and the distortion graph calculated using $J(c,\mu)=\sum_{i=1}^{m}||x_{(i)}-\mu_{c(i)}||^2$

Ex.



Inside each folder exists a png of the centroid of this cluster, a text file that has the percentage of occupancy of each class in this cluster, and a 20 sample images from this cluster where each sample is named img-class_of_sample-#order

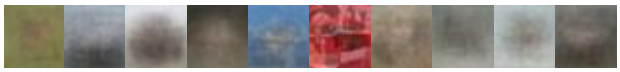


Test Runs

Samples=500 with K=10 Iterations=15:

Took 1.5 minutes.

Centroids of 10 Clusters:

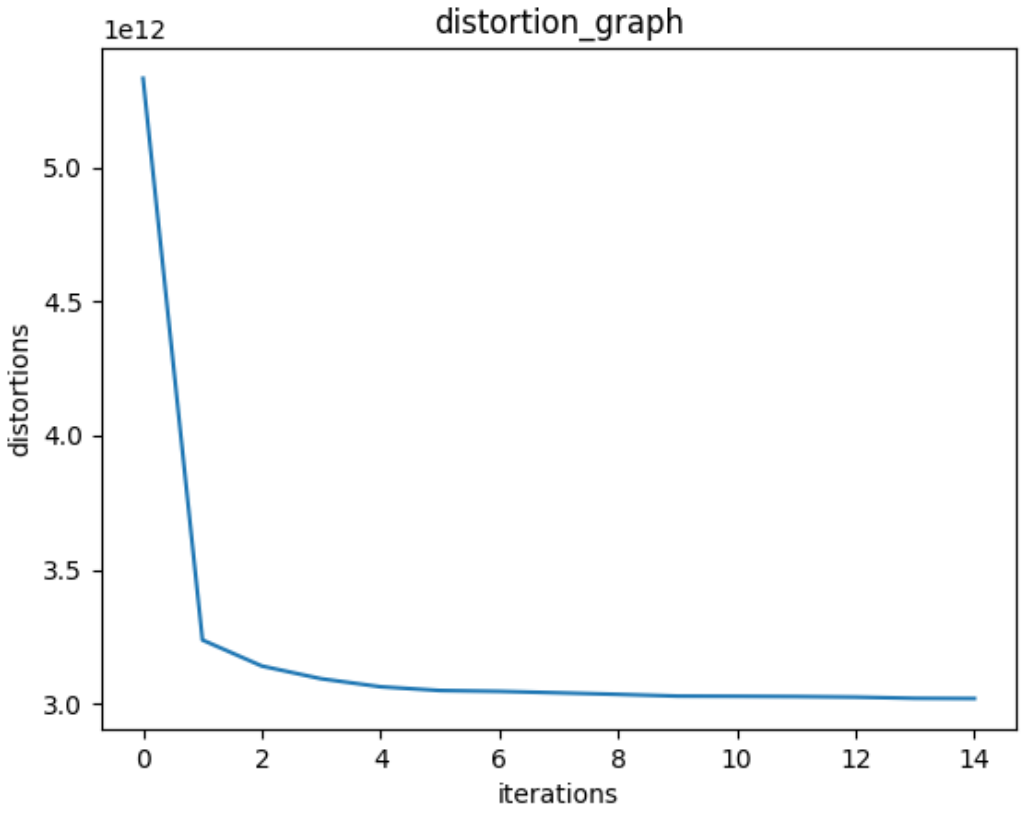


Sample from clusters:

Accuracy: 28.641%

Each cluster's accuracy is provided in a text file inside it, attached to this report.

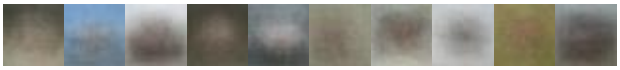
Distortion Measure:



Samples=1000 with K=10 Iterations=15:

Took 3.5 minutes.

Centroids of 10 Clusters:

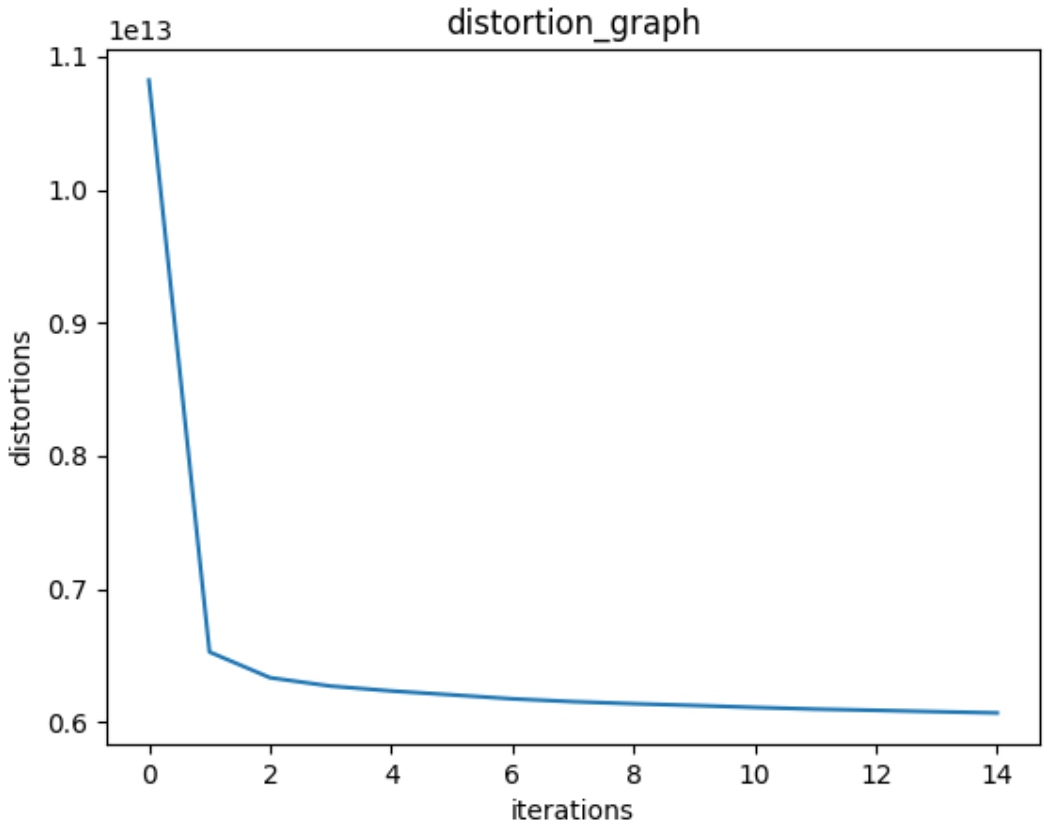


Sample from clusters:

Accuracy: 25.351%

Each cluster's accuracy is provided in a text file inside it, attached to this report.

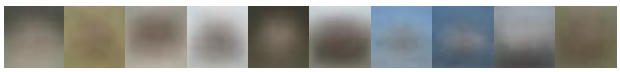
Distortion Measure:



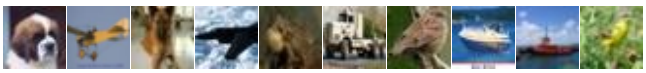
Samples=5000 with K=10 Iterations=15:

Took 15.5 minutes.

Centroids of 10 Clusters:



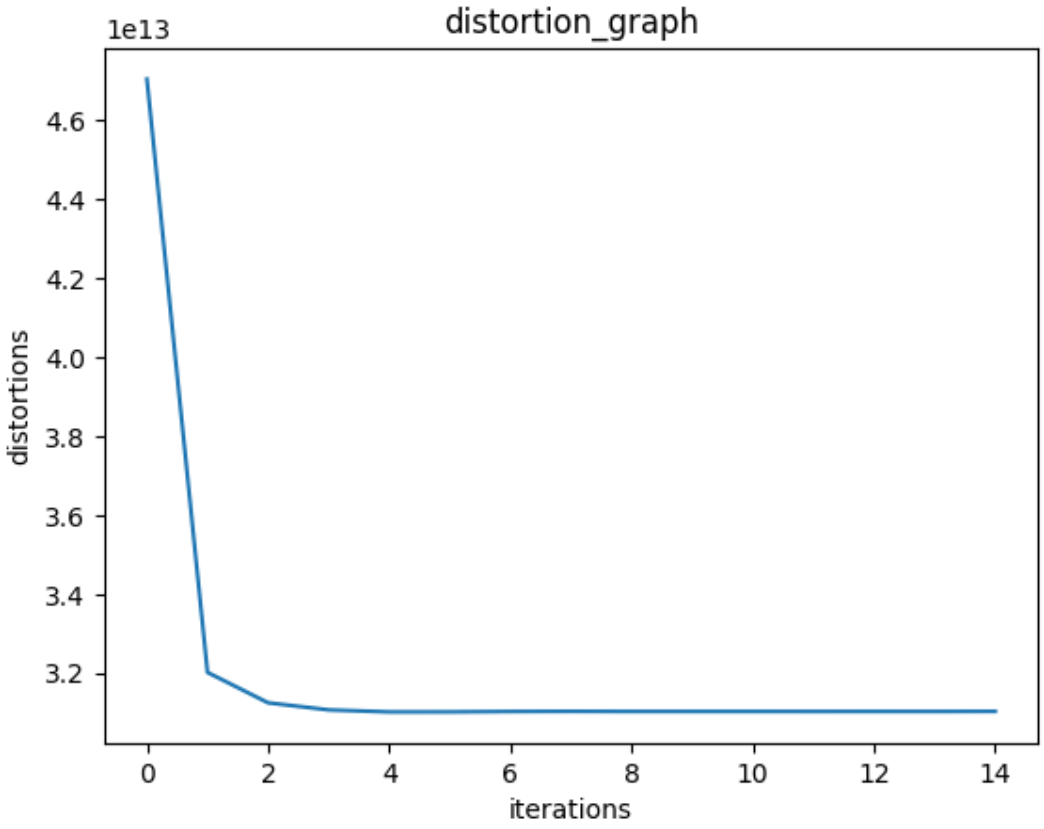
Sample from clusters:



Accuracy: 25.392%

Each cluster's accuracy is provided in a text file inside it, attached to this report.

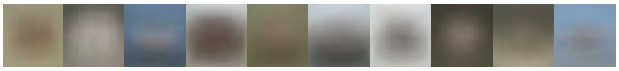
Distortion Measure:



Samples=7500 with K=10 Iterations=15:

Took 25.83 minutes.

Centroids of 10 Clusters:

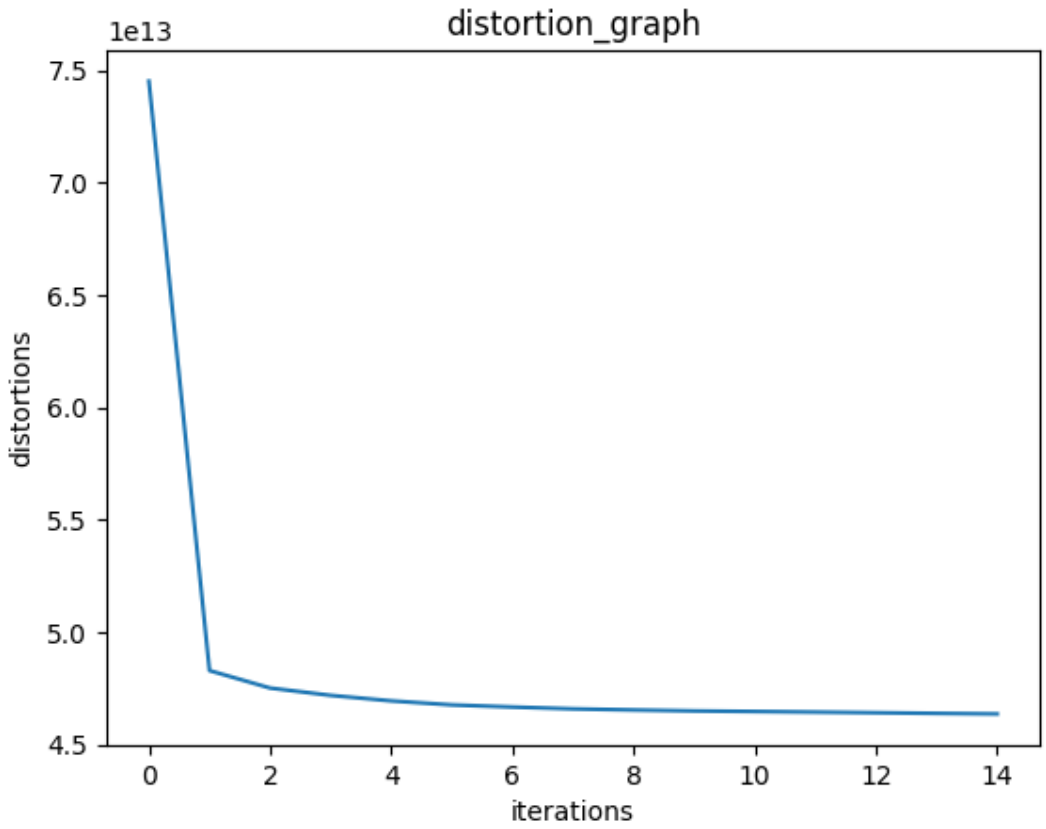


Sample from clusters:

Accuracy: 25.482%

Each cluster's accuracy is provided in a text file inside it, attached to this report.

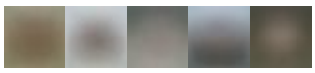
Distortion Measure:



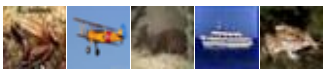
Samples=10000 with K=5 Iterations=15:

Took 20.14 minutes.

Centroids of 10 Clusters:



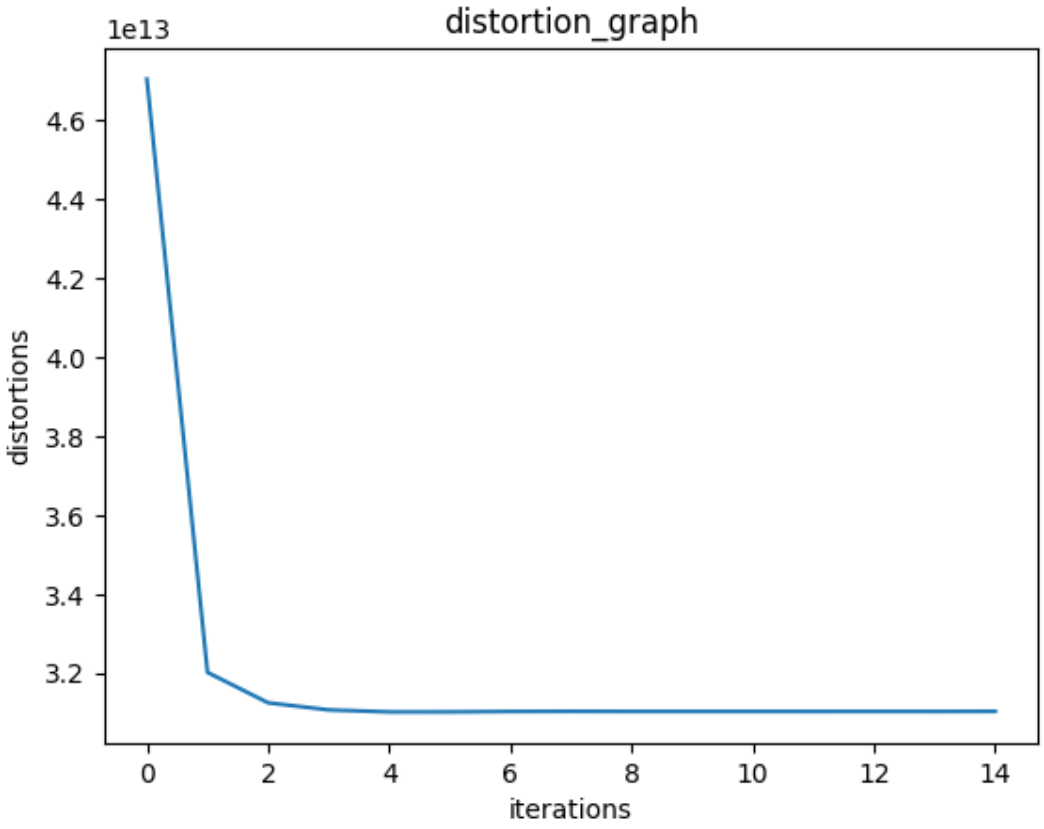
Sample from clusters:



Accuracy: 20.184%

Each cluster's accuracy is provided in a text file inside it, attached to this report.

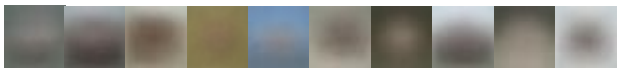
Distortion Measure:



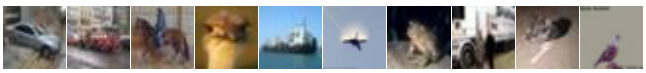
Samples=10000 with K=10 Iterations=15:

Took 64.53 minutes.

Centroids of 10 Clusters:



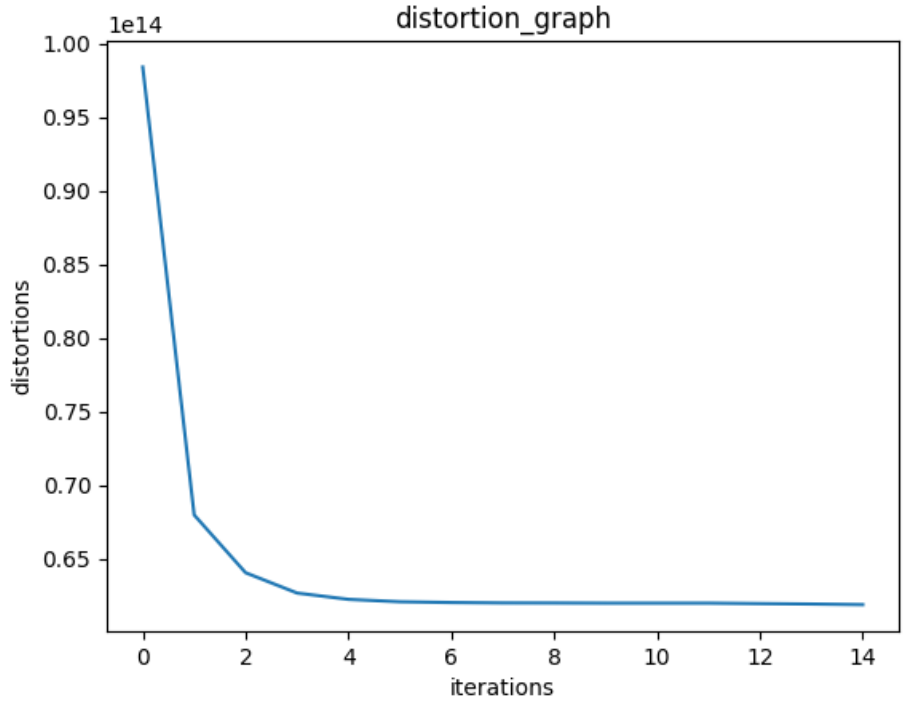
Sample from each Cluster:



Accuracy: 23.015%

Each cluster's accuracy is provided in a text file inside it attached to this report.

Distortion Measure:



Conclusion

It is clear that the distortion measure will never increase again, as the centroids, upon multiple iteration, become the center of gravity of the optimal cluster where they are at the best distance from all the points within their cluster.

As shown above, the accuracy will change slightly based different testing conditions.