Keshav Shankar
ECE 1395
Homework 9

Analyzing Number of Restarts


original


R=5


R=15


R=30

I would say that the number of restarts is not a significant factor. As you can see, the clusters are almost all the same. The only major difference is in R=5, buildings are mixed between 2 clusters, but in k=30, buildings are mostly all one cluster.

You should use more restarts when you have larger data with more dimensionality. Additionally, it could help when you have imbalanced data.

Analyzing Number of Iterations

Original

iters = 7

iters = 13

iters = 20

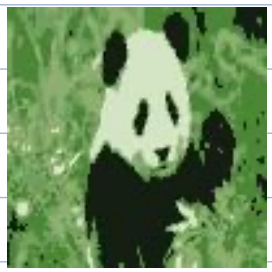
The number of iterations also seems not to matter (for this data) since all the images are identical. The only visible difference is some noisy pixels here and there, and that could have to do with iterations, since in iters = 7, there are black spots on the right, but in iters = 20, those spots are gone, because it has fully converged.

So you should use more iterations when data is tightly clustered to seperate them better.

Analyzing Number of Clusters

Original

K=3



K=5

K=7

The number of clusters immediately seems to make the most difference. In general, you can see that with more clusters, there will be more colors in an image, thus more detail. You can see in K=5, the panda is a blend of green & white, but in K=7, it is pure white.

The more clusters is usually always better for large data sets. Paired with the right iterations & restarts can produce the optimal output.

Note: For all analyses, 1 variable was analyzed while the other 2 were held constant, for sake of comparing images.

ex:

Variable	Constant	
K=3	iters=7	r=5
K=5	iters=7	r=5
K=7	iters=7	r=5