3 nov 2019

# PR  Assignment 2

1. K means clustering for linearly separable data.
2. K means clustering for non-linearly separable data.
3. Color based Image Segmentation using K means clustering.
4. Color and Position(region) based Image Segmentation using K means clustering.

## K means clustering for linearly separable data.

K means is an unsupervised learning algorithm used to classify the given sample data points to K clusters. Basic idea is to use K cluster centers, one for each. Assign each and every sample point to one of the cluster.

**Note**: No concept of training and testing data in K means.

**Given**

- Value of K - Number of clusters

**Steps**

1. Initialize K means randomly.
   Best way to initialize random mean is to find random values between minimum and maximum value of sample points.
2. Assign each sample point to the closest cluster (cluster for which euclidean distance of sample point to mean of it is minimum among all other clusters.)
3. Update the mean of each cluster after step 2.
4. Repeat the steps 2 and 3 until there is no change in assigning of sample points.

**Cost function**

## K-means Clustering

- Cost function: the sum-of-squared distances from each data point to its assigned prototype:

$$J = \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} \|x_i - \mu_k\|^2$$

- The K-means algorithm is coordinate descent on this cost function

Here, n = number of sample points

K = number of clusters

$z_{ik}$ 1 , $i^{th}$ sample point belongs to $k^{th}$ cluster.

i.e. $z_{ik}$ = 0 for other k values.

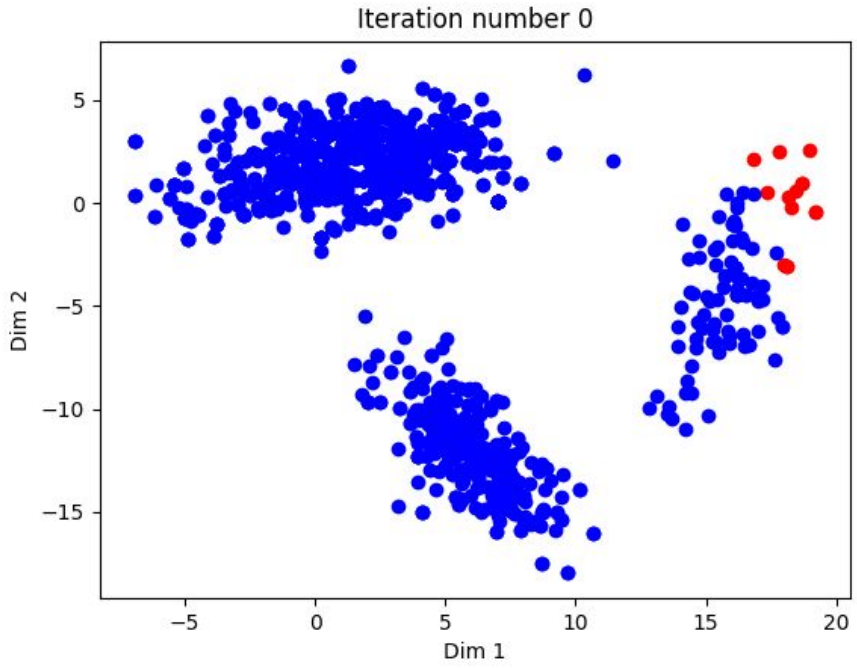
$X_i$ = $i^{th}$ sample point

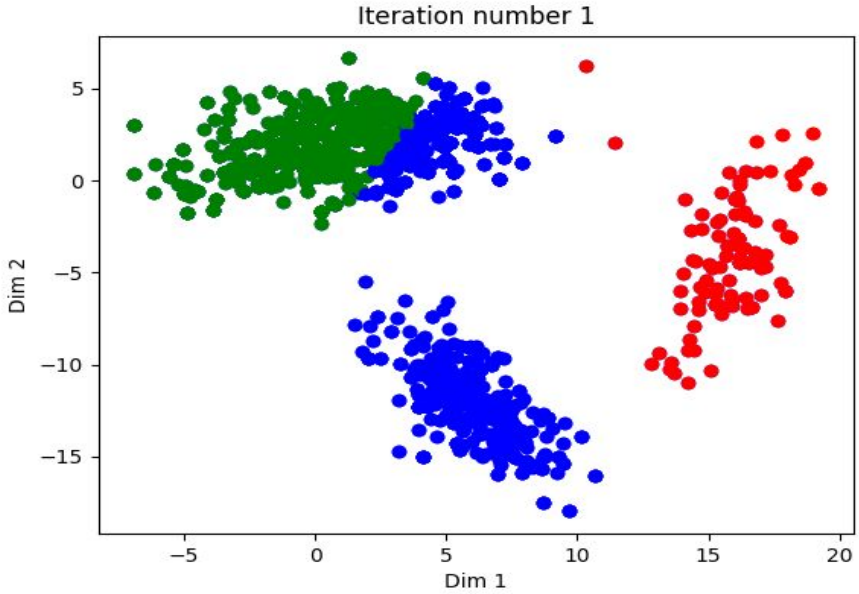
$\mu_k$ = mean of $k^{th}$ cluster

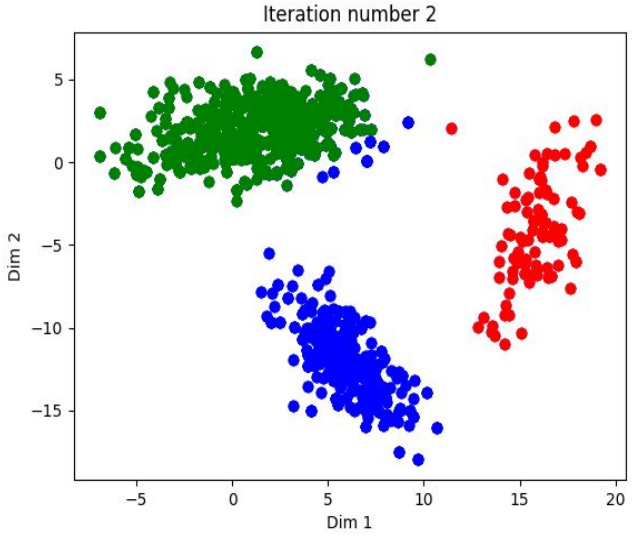
**Dataset :**

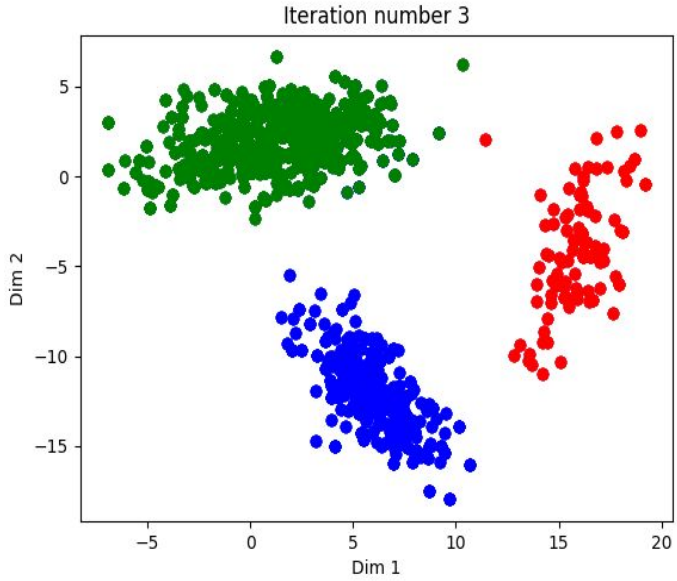
Group1 dataset was used with three classes.

## K = 3 (three classes)






Process converges after iteration 3.

Conclusion : Three clusters were formed as K = 3.

**Note** : Number of clusters totally depend on value of K.

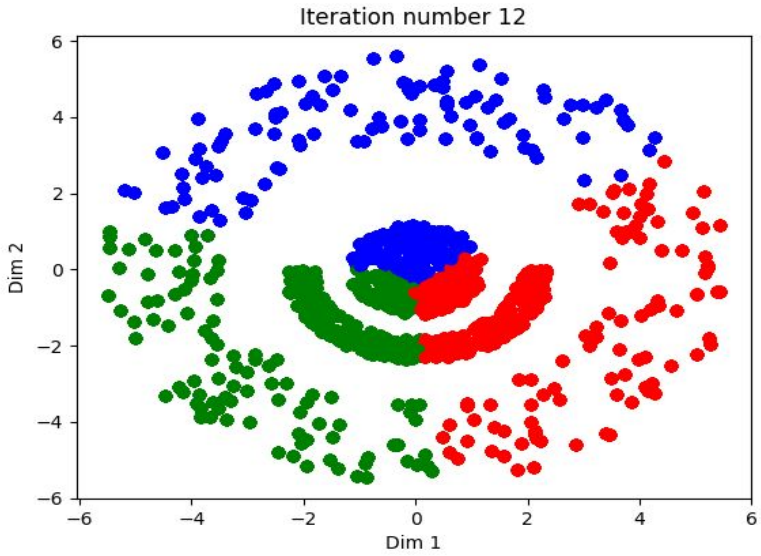
**K means clustering for non linearly separable data**

**Dataset :**

Group 3 and Group 4 dataset was used.

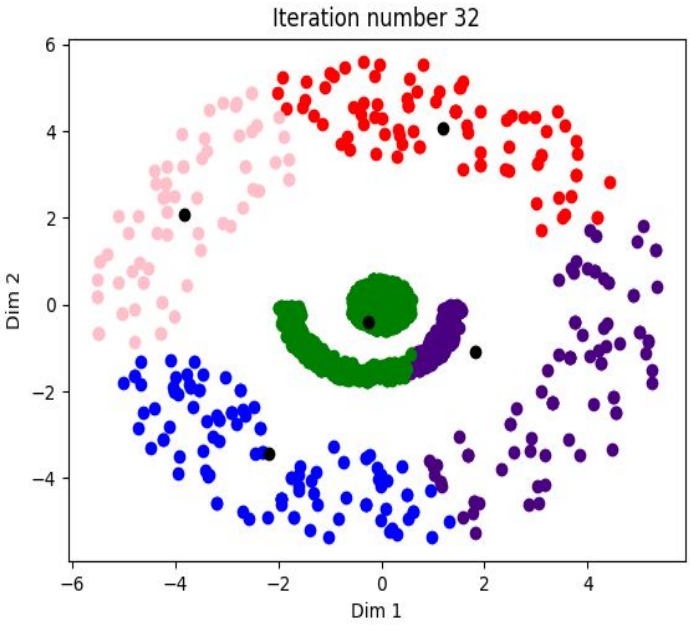
**For K = 3** :

Converges after iteration 12.




**For K = 5 :**

Converges after iteration 32.


Iteration number 32
Dim 2
Dim 1
6
4
2
0
−2
−4
−6
−4
−2
0
2
4

## Color based Image Segmentation

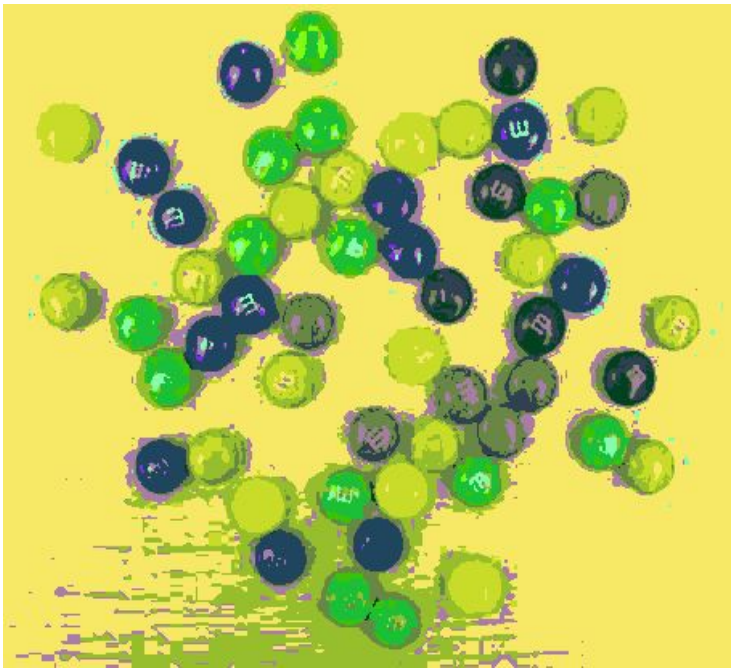
K means clustering is used to segment the given input image on the basis of color. So, basically here we are segmenting our image on the basis of colors(RGB pixel values).

**Input Image** :

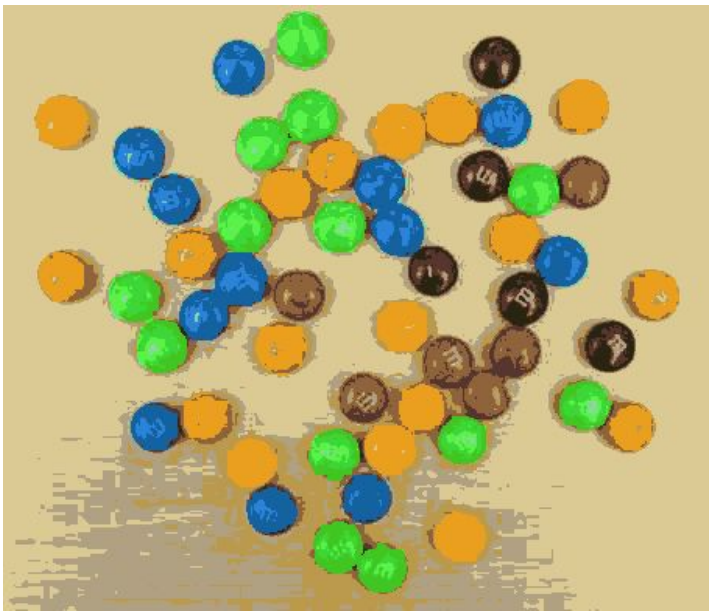
**Segmentation Image** :

**For K = 13 :**

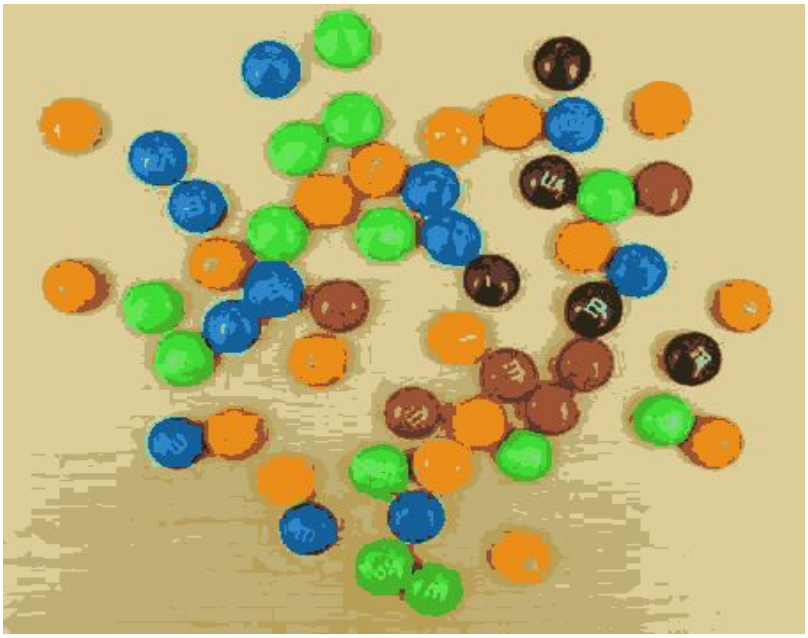
**Iteration 0** :



**Iteration 1 :**



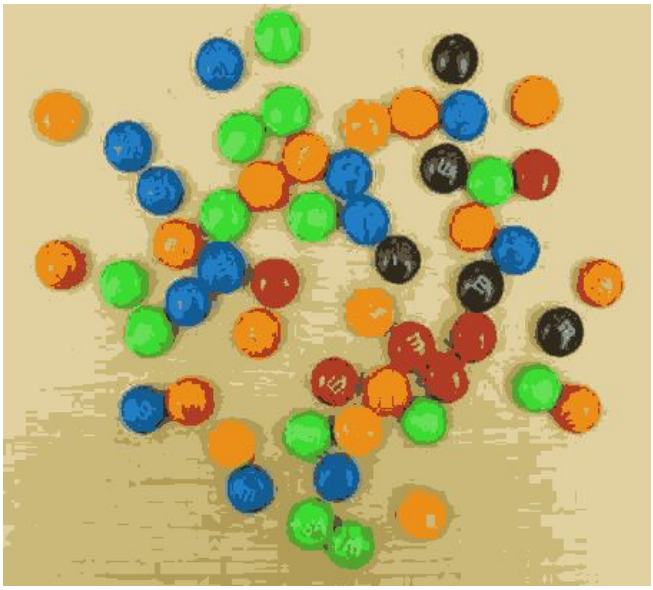
**Iteration 2 :**



**Iteration 3 :**

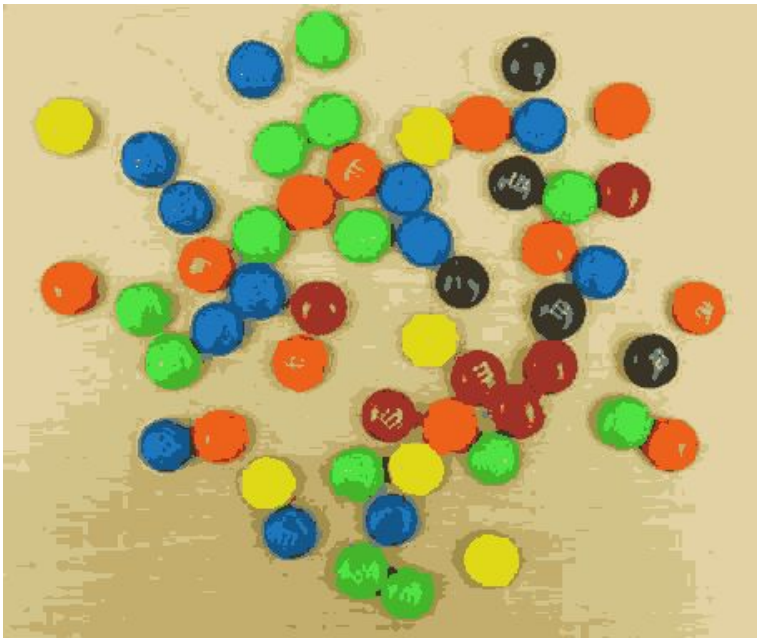
**Iteration 4 :**

**Iteration 5 :**



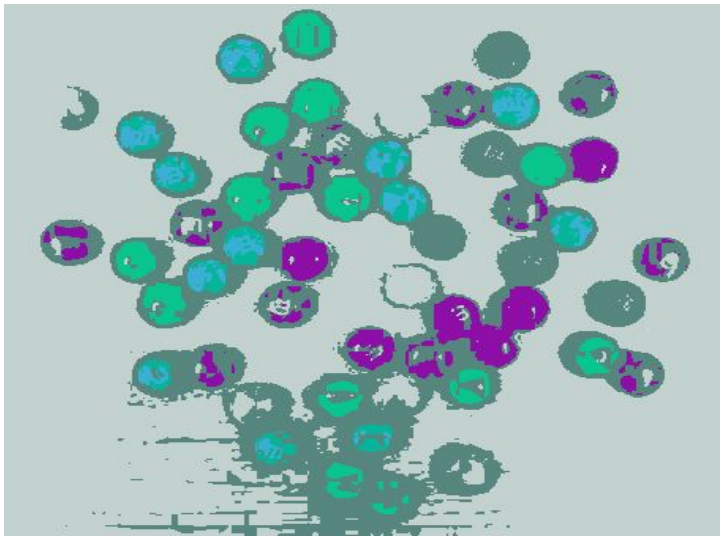
Process converges in 49th Iteration.

**Iteration 49 :**

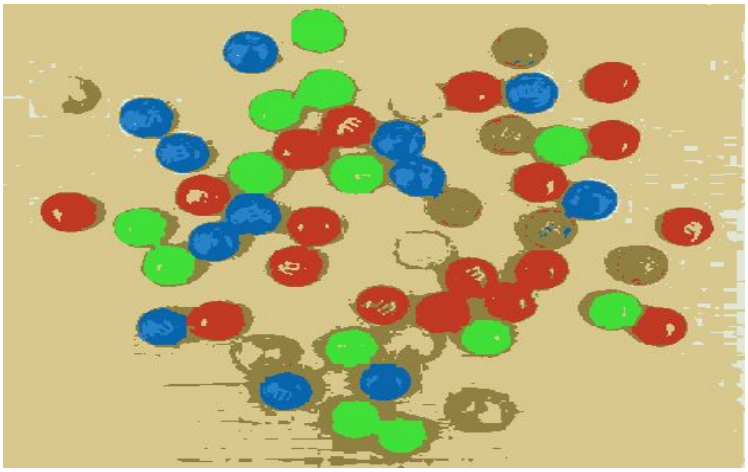


## For K = 5 :

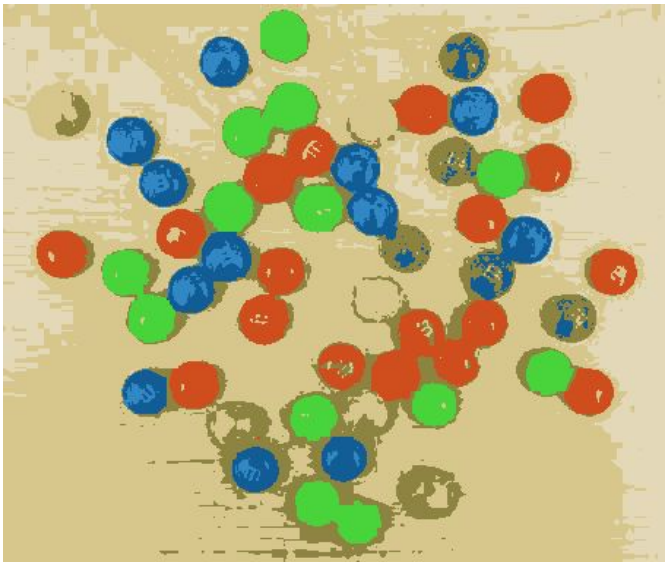
## Iteration 0 :



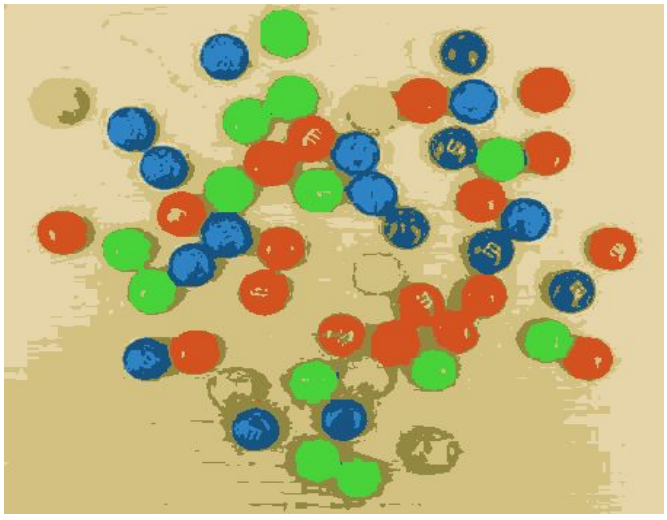
**Iteration 1 :**



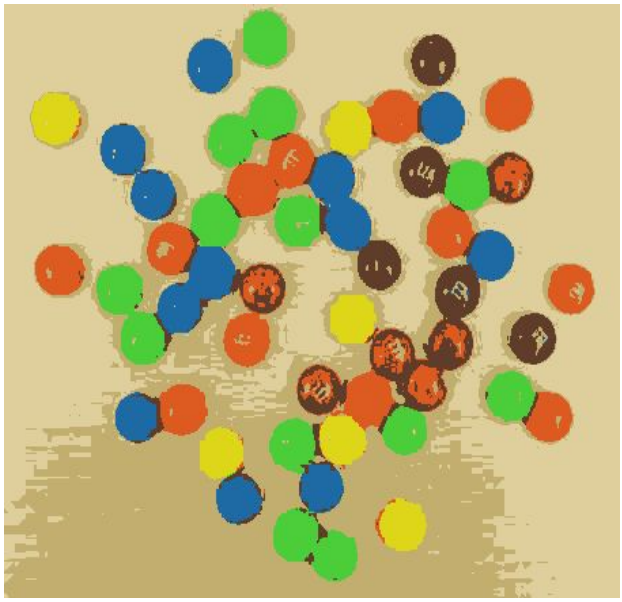
**Iteration 2 :**



**Iteration 3 :**



Process converges after in iteration 22.



## Color with location(region) based segmentation

Color and location both are used in segmentation.
Segmentation can be done by using different weights(weightage in segmentation) of location.