

```
##### Varada Gurjar , COEN 242 , HW2 4/23/2015#####
```

```
#task 1
```

```
genData <- function(N,D,K)
{
  # random uniform data of D rows and N columns as a matrix
  all_points <- matrix(runif(D*N, 1, D*N),nrow=D, ncol=N)

  centers <- matrix(nrow=D,ncol=K)

  # centroids takes random number from 1:N without replacement from it.
  centroids <- sample(1:N, K, replace=F)
  # store columns with the selected random column number in centers
  for (i in 1:K) { centers[,i] = all_points[,centroids[i]] }

  return(list(dataPoints=all_points,centers=centers))
}
```

```
#####
```

```
#task 2
```

```
closestCenters <- function(all_points, centers)
{
  # the centers and all_points are a list items. Unlist and store in matrix form
  temp <- matrix( unlist(centers),nrow=D,ncol=K ) ; centers <- temp
  temp <- matrix( unlist(all_points),nrow=D,ncol=N ) ; all_points <- temp

  # distance vector is a KxN matrix that stores  $(x_1-x_2)^2 + (y_1-y_2)^2 + (z_1-z_2)^2 \dots$  of each point
  # out of the 1:N points with 1:K centers . Initialized to all data as 1.
  distance_vector <- matrix(1,nrow=K,ncol=N)
  for(i in 1:K){ for(j in 1:N){ vector <- abs(centers[,i]-all_points[,j]) ; distance_vector[i,j] <- sum(vector*vector) } }

  # label evaluates min value out of the distances and associates each point to center
  label <- matrix(nrow=1,ncol=N)
  for(i in 1:N){ vector <- distance_vector[,i] ; for(j in 1:K){label[i] = match(min(vector),vector)} }

  new_centres <- matrix(0,nrow=D,ncol=K)
  counter <- matrix(0,nrow=1,ncol=K)

  # the counter is for count of number of points in cluster 1:K. Initialized to 0 as its values are inc by 1.
  for(i in 1:N){index <- label[i] ; new_centres[,index] <- all_points[,i]+new_centres[,index] ; counter[,index] <-
  counter[index]+1}
  for(i in 1:K){ new_centres[,i] <- new_centres[,i]/counter[,i] }

  return(list(new_centres,label))
}
```

```

#=====
#task 3

myKmeans <- function(all_points,centers,niter)
{
  # the centers and all_points are a list items. Unlist and store in matrix form.
  temp <- matrix( unlist(centers),nrow=D,ncol=K ) ; centers <- temp
  temp <- matrix( unlist(all_points),nrow=D,ncol=N ) ; all_points <- temp

  label <- matrix(nrow=1,ncol=N)
  for(i in 1:niter)
  {
    data <- closestCenters(all_points,centers)
    temp <- matrix(unlist(data[1]),nrow=D,ncol=K ) ; centers <- temp
    temp <- matrix(unlist(data[2]),nrow=1,ncol=N ) ; label <- temp
  }

  return(list(new_centers=centers,label=label))
}
#=====

```