Q2.R

mirrien

2022-04-10

```
################################################
# Question 2

manhattan <- function(p1,p2) {
  distance <- matrix(NA, nrow=dim(p1)[1], ncol=dim(p2)[1])
  for(i in 1:nrow(p2)) {
    distance[,i] = rowSums(abs(t(t(p1)-p2[i,])))
  }
  return(distance)
}

kmedian2 <- function(x,K,iters) {
  # convert df to matrix
  x = as.matrix(x)

  # randomly sample some centers, set a seed 100
  set.seed(100)
  K <- x[sample(nrow(x), K),]

  # empty lists to store outputs
  assignments <- vector(iters, mode = "list")
  locations <- vector(iters, mode = "list")

  # initialize cluster assignments using ((i-1)%K)+1

  for (i in 1:nrow(x)) {
    assignments[[i]] = ((i-1)%%K)+1
  }

  for(i in 1:iters) {
    # call manhattan distance helper function
    dists = manhattan(x,K)
    # find minimum distance
    clusters <- apply(dists,1,which.min)
    # tapply median()
    centers <- apply(x,2,tapply,clusters,median)

    # store outputs
    assignments[[i]] <- clusters
    locations[[i]] <- centers
  }
```

```
  # return outputs in list
  return(list(locations=locations[[1]], assignments = assignments[[1]]))
}

df = read.csv("parkinsons.data",row.names = 1)
result = kmedian2(df,3,1000)
print(result$locations)
```

```
##   MDVP.Fo.Hz. MDVP.Fhi.Hz. MDVP.Flo.Hz. MDVP.Jitter... MDVP.Jitter.Abs.
## 1     120.168     140.2120      97.5350       0.005275            4e-05
## 2     225.534     242.5295     202.2575       0.002735            1e-05
## 3     180.198     216.3020     109.3790       0.004600            3e-05
##   MDVP.RAP MDVP.PPQ Jitter.DDP MDVP.Shimmer MDVP.Shimmer.dB. Shimmer.APQ3
## 1 0.002685 0.003060   0.008065     0.024450           0.2305     0.013655
## 2 0.001545 0.001515   0.004635     0.016255           0.1430     0.008640
## 3 0.002370 0.002540   0.007100     0.025510           0.2550     0.014100
##   Shimmer.APQ5 MDVP.APQ Shimmer.DDA      NHR     HNR status      RPDE      DFA
## 1     0.014125 0.019525    0.040965 0.012295 22.1520      1 0.5393685 0.727867
## 2     0.009970 0.011410    0.025925 0.004760 24.8555      0 0.4294560 0.678466
## 3     0.015800 0.019090    0.042310 0.018020 20.3660      1 0.4699280 0.715121
##     spread1   spread2       D2      PPE
## 1 -5.514255 0.2318010 2.246647 0.217401
## 2 -7.162888 0.1706355 2.272559 0.095626
## 3 -5.845099 0.2180370 2.608749 0.186489
```