

# Week 10 Exercises

Aaron Palumbo

October 27, 2015

## Contents

7.6.1 (Section 7.6.7) . . . . .	1
Appendix . . . . .	2

### 7.6.1 (Section 7.6.7)

Execute the BDMO Algorithm with  $p = 3$  on the following 1-dimensional, Euclidean data:

1, 45, 80, 24, 56, 71, 17, 40, 66, 32, 48, 96, 9, 41, 75, 11, 58, 93, 28, 39, 77

The clustering algorithm is  $k$ -means with  $k = 3$ . Only the centroid of a cluster, along with its count, is needed to represent a cluster.

```
# Custom clustering library
source("clustering_library.R")

# Load in data
x <- c( 1, 45, 80, 24, 56, 71, 17,
       40, 66, 32, 48, 96, 9, 41,
       75, 11, 58, 93, 28, 39, 77)

records <- data.frame(x=x)

# Cluster parameters
p <- 3
k <- 3

# Break up data into buckets
buckets <- initBuckets(records, 3, 2)

bdmo <- lapply(buckets, function(b) {
  rec <- b[['records']]
  if (nrow(rec) > k) {
    km <- kmeans(rec, k)
    return(list(centroid = km$centers,
               count = length(km$cluster)))
  } else {
    return(list(centroid = rec,
               count = nrow(rec)))
  }
})
```

```

for (b in bdmo) {
  print("Centroid:")
  print(t(b[['centroid']]))
  print
  print(c("Count: ", b[['count']]))
  print("-----")
}

```

```

## [1] "Centroid:"
##      1      2      3
## x 1 45 80
## [1] "Count: " "3"
## [1] "-----"
## [1] "Centroid:"
##      1      2      3
## x 20.5 40 64.33333
## [1] "Count: " "6"
## [1] "-----"
## [1] "Centroid:"
##      1      2      3
## x 41 85.25 10
## [1] "Count: " "12"
## [1] "-----"

```

## Appendix

Code for initBuckets

```

initBuckets <-
function (records, ptsPerBucket, growthRate)
{
  numRecs <- nrow(records)
  numBuckets <- ((log(1 - (1 - growthRate) * numRecs/ptsPerBucket))/log(growthRate))
  cumRecords <- function(bucket) {
    ptsPerBucket * (1 - growthRate^bucket)/(1 - growthRate)
  }
  buckets <- lapply(1:numBuckets, function(b) {
    start <- ifelse(b == 1, 1, cumRecords(b - 1) + 1)
    end <- cumRecords(b)
    list(records = records[start:end, , drop = FALSE])
  })
  return(buckets)
}

```