

IS 622 Week 10 Homework

Ben Arancibia

October 31, 2015

7.6.1 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.

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

p = 3
k = 3

buckets <- function(data, k, p) {
  numberrecords <- nrow(data)
  numberbuckets <- (( log( 1 - (1-p)*numberrecords/k) )/ log(p))

  cumulativedata <- function(bucket) {
    k * (1 - p**bucket) / (1 - p)
  }

  buckets <- lapply(1:numberbuckets, function(b) {
    start <- ifelse(b == 1, 1, cumulativedata(b - 1) + 1)
    end <- cumulativedata(b)
    list(data=data[start:end, , drop=FALSE])
  })
  return(buckets)
}

buckets(x, 3, 3)

## [[1]]
## [[1]]$data
##      x
## 1  1
## 2 45
## 3 80
##
##
## [[2]]
## [[2]]$data
##      x
## 4  24
## 5  56
## 6  71
## 7  17
## 8  40
## 9  66
```

10 32
11 48
12 96