week1

Rajesh Kalakoti

2023-08-03

* Packages
  + [devtools](https://www.r-project.org/nosvn/pandoc/devtools.html)
  + [tidyverse](https://www.tidyverse.org/packages/)
    - sub-sub-item 1

## Clustering

Given a clustering , we need some scoring function that evaluates its quality or goodness. This sum of squared errors scoring function is defined as:

The goal is to find the clustering that minimizes:

K-means employs a greedy iterative approach to find a clustering that minimizes loss function.

## Algorithm 13.1: K-means Algorithm

**K-means** (*D*, *k*, *ε*): 1. Initialize . Randomly initialize centroids: . 3. **repeat** 4. . **// Cluster Assignment Step** 5. **foreach** **do** 6. . **// Assign to closest centroid** 7. . **// Centroid Update Step** 8. **foreach** **to** **do** 9. . **until** .

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

?entropy

## No documentation for 'entropy' in specified packages and libraries:  
## you could try '??entropy'