Robert Morson

Assignment 5

- The best method of the linkage clustering is the average linkage.
- Choosing from k = 2,3,4,5,6
    - K = 2 gives the highest silhouette value
- Chosen numbers of clusters = 2

Cluster structures:

Cluster 1 – 3 cereals, very small

- 100%_Bran
- All – Bran
- All – Bran_with_Extra_Fiber

This is the healthy cluster. High in fiber, low sugar, low fat, lower calories

Cluster 2 – 71 remaining cereals

Sweeter, more processed, less fiber

▪ Stability (split-half check)

- Randomly split cereals into two halves A and B.
- Clustered A (average linkage, k = 2), computed centroids, assigned B by nearest centroid.
- Compared these assignments to the clusters from using all data.

Result:

- About 92% of cereals in B get the same cluster label as in the full-data clustering.

Conclusion: Clustering is reasonably stable.

▪ Healthy cereals cluster & normalization

- For school cafeterias, choose cereals from Cluster 1 (healthiest: high fiber, low sugar/fat).
- Yes, data should be normalized (z-scores) before clustering, because variables (calories, sodium, fiber, etc.) are on different scales, and we don't want any one variable to dominate the distance.