COMP257 Sec.402 Unsupervised & Reinforcement Learning

Assignment 3: Hierarchical Clustering

Name: Damien Liscio

Student #: 301237966

Due: Sunday, October 13th, 2024

5. Discuss any discrepancies observed between 4(a), 4(b), or 4(c).

There were some discrepancies observed between the different similarities, but also some commonalities. The first two similarity measures, Euclidean Distance & Minkowski distance yield the same silhouette scores. This is due to the fact the Euclidean distance is the same as the Minkowski distance when the order of the Minkowski distance is 2 (p=2), which it is set to in 4(b) to yield the highest silhouette scores. Both differ from Cosine Similarity, which had a slightly lower silhouette score than the other two. This makes sense as while the first two use distance to cluster, the Cosine Similarity uses the angle between vectors to cluster data points, resulting in different (and this case lower) silhouette scores.