Chapter 2: Introduction to Unsupervised Learning Answer Key

Activity 1 Solutions:

Please refer to the provided Jupyter notebook for solution code.

Activity 2 Solutions:

Please refer to the provided Jupyter notebook for solution code.

Practice Question Solutions:

1. The two approaches to Hierarchical Clustering are Agglomerative and Degenerative
   1. True
   2. **False**
2. Which of these linkage criteria can be used to determine distances between clusters?
   1. Complete
   2. Single
   3. Centroid
   4. **All of the above**
3. Hierarchies are always the most efficient way to convey information about a data set
   1. True
   2. **False**
4. How is Hierarchical Clustering a better approach to clustering than K-Means?
   1. It is faster than K-Means
   2. **You don’t need to determine number of clusters a priori**
   3. It handles simple data sets in a more efficient manner
   4. All of the above
5. How is K-Means a better approach to clustering than Hierarchical Clustering?
   1. **It is faster than Hierarchical Clustering**
   2. It is slower but more accurate
   3. You don’t need to determine what K is a priori
   4. K-Means is much more complex but more accurate
6. A hierarchy shows the relationships from parent nodes to their children nodes
   1. **True**
   2. False
7. In a hierarchy, children nodes can also themselves be parents to other children nodes
   1. **True**
   2. False
8. Why are dendrograms helpful in determining clusters?
   1. It shows how correlated features in the data set are
   2. **They give an idea how clusters could be formed with which members**
   3. If they fail, it’s a sign that the data is too complex for hierarchical clustering
   4. They aren’t helpful
9. Agglomerative clustering adopts the tops-down approach, while Divisive adopts the bottoms-up approach.
   1. True
   2. **False**
10. Hierarchical Clustering in sci-kit learn is more modular and should always be preferred over scipy implementations.
    1. True
    2. **False**