Chapter 3: 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 hyperparameters that primarily drive performance in DBSCAN are neighborhood radius size and minimum number of points threshold.
   1. **True**
   2. False
2. Which of these clustering methods require specifying the number of clusters prior to runtime?
   1. **K-Means**
   2. Hierarchical Clustering
   3. DBSCAN
   4. A & B
3. DBSCAN uses solely the neighborhood measure and doesn’t use any distance functions.
   1. True
   2. **False**
4. Which of these clustering methods only need to cycle through all of your data points just one time?
   1. K-Means
   2. Hierarchical Clustering
   3. **DBSCAN**
   4. A & B
5. DBSCAN focuses on capturing what notion when it comes to approximating neighbors?
   1. Dilution
   2. **Density**
   3. Differential
   4. None of the Above
6. The neighborhood in a normal implementation of DBSCAN can be visualized as a circle
   1. **True**
   2. False
7. The end result of DBSCAN can have many data points assigned to no cluster
   1. **True**
   2. False
8. Typically, in what scenario will DBSCAN excel as a clustering method?
   1. Discrete blobs of data points
   2. Sandwiched data points where linear relationships are difficult to draw
   3. When there is a lot of noise in your data
   4. **B & C**
9. You should always default to DBSCAN as the best clustering method and then evaluate K-Means and Hierarchical Clustering only if DBSCAN fails
   1. True
   2. **False**
10. Neighborhood radius size and minimum number of points will not change as a result of how your data looks – good combinations of the hyperparameters will always be good combinations.
    1. True
    2. **False**