Clustering

1. True or False: Most genes in DeRisi's 6400-gene dataset are likely not involved in the diauxic shift.

   True

   False

2. True or False: The fact that only 13% of genes in S. cerevisiae are duplicated does not rule out the possibility of a whole-genome duplication.

   True

   False

3. True or False: In practice, the number of clusters of biological data is typically known in advance.

   True

   False

4. True or False: Center placement in the k-Centers Clustering Problem is generally more sensitive to outliers than center placement in the k-Means Clustering Problem.

   True

   False

5. How many ways are there to cluster 5 points into two clusters?

   15

6. True or False: How we choose initial centers for the Lloyd algorithm can make a difference.

    True

    False

7. Compute MaxDistance(Data, Centers) for the following Data and Centers:

    Data: (2, 8), (2, 5), (6, 9), (7, 5), (5, 2)   Centers: (3, 5), (5, 4)

    5

8. Compute *Distortion*(*Data, Centers*) for the following *Data* and *Centers*:

    Data: (2, 8), (2, 5), (6, 9), (7, 5), (5, 2)   Centers: (3, 5), (5, 4)

    9

9. Give the center of gravity of the following (three-dimensional) data points. Enter your answer in the form (x, y, z). (Please note the space between the coordinates.)

    (17, 0, -4), (3, 14, 23), (9, 7, 16), (7, 3, 5)

    (9, 6, 10)

Advanced Clustering

1. Say that a coin is weighted so that it produces a single heads with probability $\theta = 0.7$. Compute $\Pr(HTHHT|\theta)$. Give your answer to three decimal places.

   0.031

2. Given the following Data and Centers, compute HiddenMatrix2,4 (i.e., the responsibility of the second center for the fourth datapoint) using the partition function with stiffness parameter equal to 1. Give your answer to three decimal places.
   Data: (2,8), (2,5), (6,9), (7,5), (5,2)
   Centers: (3,5), (5,4)

   Not 1

3. Say we have the following Data and HiddenMatrix:
   Data: (2,6), (4,9), (5,7), (6,5), (8,3)
   HiddenMatrix:
   0.6 0.1 0.8 0.5 0.7
   0.4 0.9 0.2 0.5 0.3
   Compute the weighted center of gravity corresponding to the first row of HiddenMatrix. Enter the coordinates of the weighted center of gravity as a pair space-separated numbers rounded to three decimal places.

   5.259 5.444

4. Below is a tree used by **HierarchicalClustering**. Which of the following clusters can be inferred from this tree? (Select all that apply.)

{1, 7, 8}, {2}, {3}, {4}, {5}, {6}, {9}, {10}

{1, 7, 8}, {2}, {3, 5, 6}, {4}, {9}, {10}

{1, 7, 8}, {2}, {3, 6}, {4}, {5}, {9}, {10}

{1, 7}, {2, 10}, {3, 6}, {4}, {5}, {8}, {9}
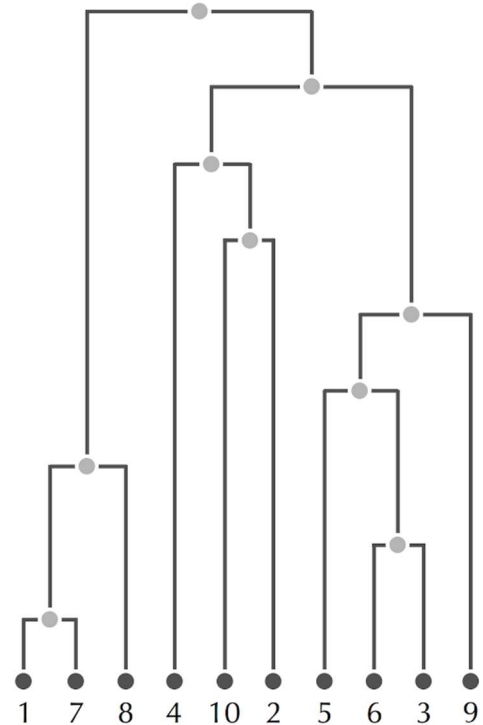
{1}, {2, 10}, {3}, {4}, {5}, {6}, {7}, {8}, {9}

{1, 2, 3, 4, 5, 6}, {7, 8, 9, 10}

{1, 7, 8}, {2, 10}, {3, 5, 6, 9}, {4}

{1, 7, 8}, {2}, {3, 5, 6, 9}, {4}, {10}

{1, 7}, {2}, {3}, {4}, {5}, {6}, {8}, {9}, {10}

5. Below is a distance matrix D. If $C_1 = \{i, l\}$ and $C_2 = \{j, k\}$, compute $D_{avg}(C_1, C_2)$.

```
  i  j  k  l
i 0 20  9 11
j 20 0 17 11
k 9 17  0  8
l 11 11  8  0
```

12

6. Below is a distance matrix D. If $C_1 = \{i, l\}$ and $C_2 = \{j, k\}$, compute $D_{min}(C_1, C_2)$.

```
  i  j  k  l
i 0 20  9 11
j 20 0 17 11
k 9 17  0  8
l 11 11  8  0
```

8

Population Genetics

1. True or False: The out of Africa hypothesis relates to a migration of Homo erectus 2 million years ago.

   True

   False

2. True or False: Modern Africans are more genetically diverse than non-Africans.

   True

   False

3. True or False: The mitochondrial genome consists of mostly "junk" DNA serving no real purpose to the mitochondrion.

   True

   False

4. The oldest Homo sapiens fossils ever discovered are called the ___.

   Omo remains

5. ___ were a first wave of European human settlers who descended directly from *Homo erectus*.

   Neanderthals

6. The exchange of genetic material between populations is called ___.

   gene flow

7. Compute Diff(s, t) for the SNPs s = (0, 0, 1, 1, 0, 0, 1, 0) and t = (1, 1, 0, 0, 1, 1, 1, 1).
    Express your answer as a decimal between 0 and 1 to three decimal places.

    Diff(s, t) = 10/12 = **0.833**

8. How many possible columns of length 5 are compatible with the column (1, 0, 1, 0, 1)?

    14