AMAT583 (8434) Midterm II

Name ..... Score: .../18

Problem 1. (3 points) What is the edit distance between

a. $AAAA$ and BBB?

b. $ABABAB$ and $BABABA$?

c. 000111 and 11100?

Explain your answers.

Solution:

a. 4. $AAAA \to AAA \to BAA \to BBA \to BBB$.

b. 2. $ABABAB \to BABAB \to BABABA$.

c. 5. $000111 \to 00111 \to 0111 \to 111 \to 1110 \to 11100$.

Problem 2. (3 points) Compute the Wasserstein distance between $(4, 2, 1)$ and $(2, 1, 4)$. (You can also view these as functions $f, g\colon \{1, 2, 3, 4\} \to [0, \infty)$ given by $f(1) = 4, f(2) = 2, f(3) = 1$ and $g(1) = 2, g(2) = 1, g(3) = 4$.)

Solution: An optimal transportation plan is given by $\begin{bmatrix} 2 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$. The cost is $2 \cdot 2 + 1 \cdot 1$, since there is a 2 two steps off the diagonal and a 1 one step off the diagonal.

Problem 3. (3 points) Let $X = \{1, 2, 3, 4\} \subseteq \mathbb{R}$. Find the k-means clustering of X with $k = 1, 2, 3, 4$. If any of these have several correct solutions, write all of them.

Solution:

$k = 1$: $\{\{1, 2, 3, 4\}\}$

$k = 2$: $\{\{1, 2\}, \{3, 4\}\}$

$k = 3$: $\{\{1\}, \{2\}, \{3, 4\}\}$ and $\{\{1\}, \{2, 3\}, \{4\}\}$ and $\{\{1, 2\}, \{3\}, \{4\}\}$ are all solutions.

$k = 4$: $\{\{1\}, \{2\}, \{3\}, \{4\}\}$

Problem 4. (3 points) Draw a dendrogram that has the barcode $\{[1, \infty), [3, 6), [2, 5), [1, 4)\}$.

Solution: See Fig. 1a.

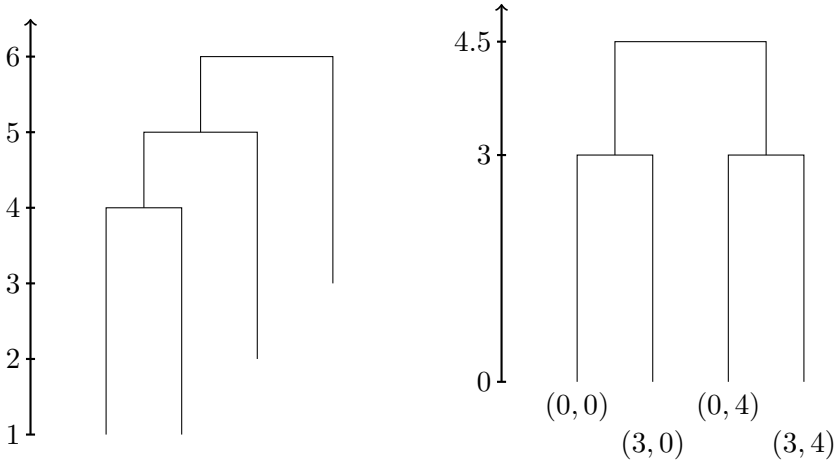
Problem 5. (3 points) Find the average linkage dendrogram of $\{(0, 0), (3, 0), (0, 4), (3, 4)\}$ equipped with the Euclidean metric.

Solution: See Fig. 1b. The key part of the calculations is $d_2((0, 0), (3, 4)) = d_2((3, 0), (0, 4)) = 5$, which gives

$$\delta(\{(0, 0), (3, 0)\}, \{(0, 4), (3, 4)\}) = \frac{1}{2 \cdot 2}(4 + 4 + 5 + 5) = 4.5.$$

Problem 6. (3 points) Find a set of points $X = \{a, b, c, d\} \subseteq \mathbb{R}$ whose single linkage clustering dendrogram is the one in Fig. 2.

Solution: A solution is $a = 0$, $b = 1$, $c = 4$, $d = 7$, which gives $X = \{0, 1, 4, 7\}$.



(a) Solution to Problem 4.

(b) Solution to Problem 5.

Figure 1

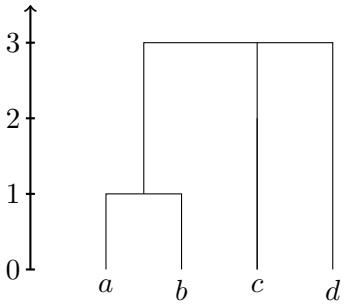


Figure 2: A dendrogram.