Practice Questions.

1. A question on multiple alignment:
Suppose S_1, S_2, S_3, S_4, and S_5 are strings and that you have the following data for pairwise alignment using *distances*.

	S_1	**S_2**	**S_3**	**S_4**	**S_5**
S_1	0	8	1	6	3
S_2		0	4	7	4
S_3			0	9	2
S_4				0	8
S_5					0

Recall the center-star method for finding an approximation to the optimal multiple alignment.

(a) With only the available information, which of the above strings would be most appropriate to use as the center? Why? Note that we are using *distance* and not *similarity*.

(b) Suppose we want to find a multiple alignment for the four strings T_1, T_2, T_3, and T_4 using the center star method. If T_1 is the center and the following pairwise alignments are known, determine the resulting multiple alignment using the "once a gap—always a gap" rule.

T_1: `BAC-CBDD` **T_2:** `-ACBABDD` **T_1:** `-BACCBD-D` **T_3:** `DBCCC-DAD` **T_1:** `B-AC-CBDD` **T_4:** `BCAADC-DD`	**T_2:** `ACBAB-D-D` **T_3:** `-DBCCCDAD` **T_2:** `ACBA-BDD` **T_4:** `BCAADCDD`	**T_3:** `DBCCCDAD-` **T_4:** `-BCAADCDD`

2. A question on pairwise structure comparison:
Consider the following two point sets.

Point Set A	**Point Set B**
(0, 0, 0)	(1, 7, 15)
(3, 4, 0)	(5, 10, 15)
(5, 0, 12)	(1, 2, 3)
(12, 3, 4)	(4, 19, 19)

(a) Compute the RMSD (Root Mean Square Distance) between the two point sets; assume that the point sets are fixed in position (i.e., no rotation, no translation). Your answer should contain one or more square roots. You do not have to do all the arithmetic as long as we can tell that you are using the formula correctly.

(b) Compute the distance matrix for each point set. It is all right if your answer includes some square roots. Do the distance matrices indicate that the point sets match or not? Why?

3. Explain qualitatively why optimal alignment in threading using a pair potential for scoring (a sum of interactions between pairs of amino acids) cannot be found efficiently.

4. When comparing gene expression profiles

a) what is the difference between the Euclidean metric and the Pearson correlation? What does the Euclidean metric emphasize, and what does the Pearson correlation emphasize?

b) What is the advantage of applying cluster analysis over pairwise analysis?

c) Explain the notions of an eigengene and an eigenarray.

How are they derived?

- Say the first eigengene (with the largest eigenvalue) has large positive value in its first component, a large negative value in the second component and zero in all other components.

What is the meaning of this pattern?

- Say the first eigenarray has large positive values for components i1,i2 and i3 and negative values for j1 and j2, and 0 otherwise. What is the meaning of this pattern?