INF281 Exercise 06 solutions

1. **SP score**

   The SP (sum-of-pairs) score is used to evaluate multiple sequence alignments.

   $$S(\mathcal{A}) = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} S(\bar{s}^i, \bar{s}^j)$$

   Use the simple scoring scheme and the alignment $\mathcal{A}$ to answer the following questions.

   Alignment $\mathcal{A}$:

   ```
   Seq1: A-G
   Seq2: GCG
   Seq3: G-T
   ```

   Scoring scheme:
   $R_{ab} = 1$ for a = b
   $R_{ab} = 0$ for a ≠ b
   $g = 1$

   N.B. The score of a column with two blanks in a pairwise alignment should be 0.

   (a) Calculate the pairwise score $S(\bar{s}^1, \bar{s}^2)$.

   Solution: 0

   (b) Calculate the pairwise score $S(\bar{s}^1, \bar{s}^3)$.

   Solution: 0

   (c) Calculate the pairwise score $S(\bar{s}^2, \bar{s}^3)$.
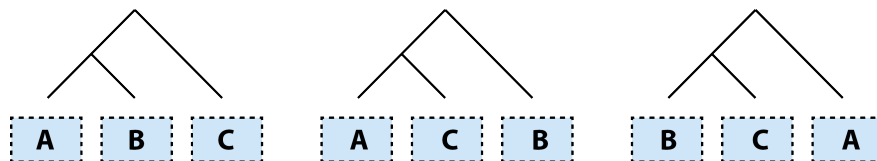
   Solution: 0

   (d) Calculate the SP score $S(\mathcal{A})$.

   Solution: $0 + 0 + 0 = 0$

2. **Tree topology**

   A rooted phylogenetic tree can have three topologically different trees when m is 3.

   (a) Fill the labels A, B, or C to satisfy three topologically distinct trees.

3. UPGMA

UPGMA is an unweighted version of PGMA (pair-group method using arithmetic mean) for reconstructing a phylogenetic tree. Pairwise sequence alignments are used to calculate the distances among four sequences A, B, C, and D.

	A	B	C	D
A	0	2	7	7
B		0	5	9
C			0	8
D				0

Below are two examples of the distance calculation that can be used for UPGMA.

$$d_{(\alpha\beta),\gamma} = \frac{d_{\alpha,\gamma} + d_{\beta,\gamma}}{2}, \quad d_{(\alpha\beta\gamma),\delta} = \frac{d_{\alpha,\delta} + d_{\beta,\gamma} + d_{\delta,\gamma}}{3}$$
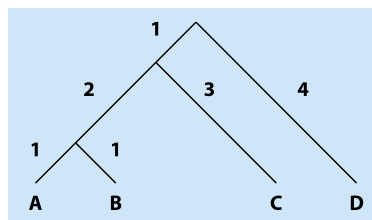
(a) Identify the first internal node and update the distance matrix.

	(AB)	C	D
(AB)	0	6	8
C		0	8
D			0

(b) Identify the second internal node and update the distance matrix accordingly.

	(ABC)	D
(ABC)	0	8
D		0

(c) Reconstrut a rooted tree from the calcualted distances.

(d) Reconstruct a rooted tree by using UPGMA.

	A	B	C	D
A	0	2	6	8
B		0	6	8
C			0	8
D				0