10 Exercise solutions – Progressing alignment

1. **Linkage clustering method for progressive alignment**

   Select two alignments from the three alignments, $\mathcal{A}^1 = s^1$, $\mathcal{A}^2 = s^2, s^3$, and $\mathcal{A}^3 = s^4, s^5$, for clustering.

   Pairwise scores

	s^1	s^2	s^3	s^4	s^5
s^1	-	2	3	1	6
s^2		-	-	4	5
s^3			-	3	4
s^4				-	-
s^5					-

   (a) Use the average linkage.

   Solution: $\mathcal{A}^2$ and $\mathcal{A}^3$: $S(\mathcal{A}^1, \mathcal{A}^2) = 2.5$, $S(\mathcal{A}^1, \mathcal{A}^3) = 3.5$, $S(\mathcal{A}^2, \mathcal{A}^3) = 4$

   (b) Use the maximum linkage.

   Solution: $\mathcal{A}^1$ and $\mathcal{A}^3$: $S(\mathcal{A}^1, \mathcal{A}^2) = 3$, $S(\mathcal{A}^1, \mathcal{A}^3) = 6$, $S(\mathcal{A}^2, \mathcal{A}^3) = 5$

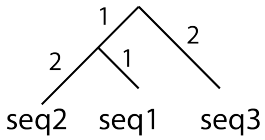
   (c) Use the minimum linkage.

   Solution: $\mathcal{A}^2$ and $\mathcal{A}^3$: $S(\mathcal{A}^1, \mathcal{A}^2) = 2$, $S(\mathcal{A}^1, \mathcal{A}^3) = 1$, $S(\mathcal{A}^2, \mathcal{A}^3) = 3$

2. **Linear progressive alignment**

Construct an MSA from seq1, seq2, seq3 and a phylogenetic tree by using the progressive alignment method specified below.

seq1: AGCT
seq2: ACT
seq3: AGCAT



- Clustering: Linear clustering
- Aligning method: Pair-guided alignment
- Aligning order: Use the specified tree
- Pairwise DP: Global alignment with linear gap penalty
- DP scoring scheme: match (10), mismatch (-5), gap penalty (10)

(a) What is the aligning order that can be defined by the given tree?

Solution:

```
1: Seq2 & Seq1
2: Seq1 & Seq3
```

(b) Solve the first pairwise alignment.

Solution:

```
Seq2: A-CT
Seq1: AGCT
```

(c) Solve the second pairwise alignment.

Solution:

```
Seq1: AGC-T
Seq3: AGCAT
```

(d) Find the optimal MSA by combining the first and the second alignments.

Solution:

```
Seq1: AGC-T
Seq2: A-C-T
Seq3: AGCAT
```

(e) What is the SP score of the optimal MSA?

Solution:

```
Seq1 & Seq2: 20
Seq1 & Seq3: 30
Seq2 & Seq3: 10

SP: 60
```