Homework 3 - Theory Questions

1) When we have two very dissimilar sequences and try to calculate the K80 distance, we first get a larger than usual transition proportion (S) and transversion proportion (V).

$$\text{K80 distance: } \hat{d} = -\tfrac{1}{2}\log(1 - 2S - V) - \tfrac{1}{4}\log(1 - 2V)$$

When filling out the above equation with S and V it is very likely that the terms inside the log() will result in negative numbers, making the entire calculation impossible, since we cannot take the log of a negative number.

2) Whichever minimal entry is taken first will influence the outcome of the tree. The second and third tips will be exchanged with each other depending on which minimal distance is taken first.
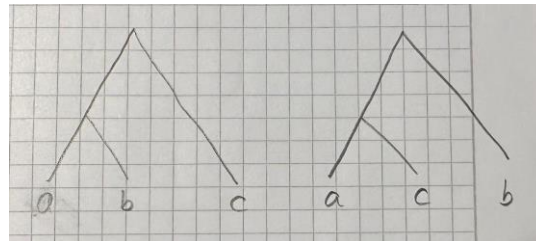
3) We would expect the distances to be the same if time (the molecular clock) correlated exactly with genetic distance, assuming the same evolutionary speed on all lineages. But sadly, this is never the case. The UPGMA algorithm produces an ultrametric tree in which the distances from the root to every branch tip are the same. Also, the UPGMA algorithm can only make rooted trees. It cannot re-root.

4) The least squares method could help us in this case to make unrooted trees. And also non-ultrametric trees.