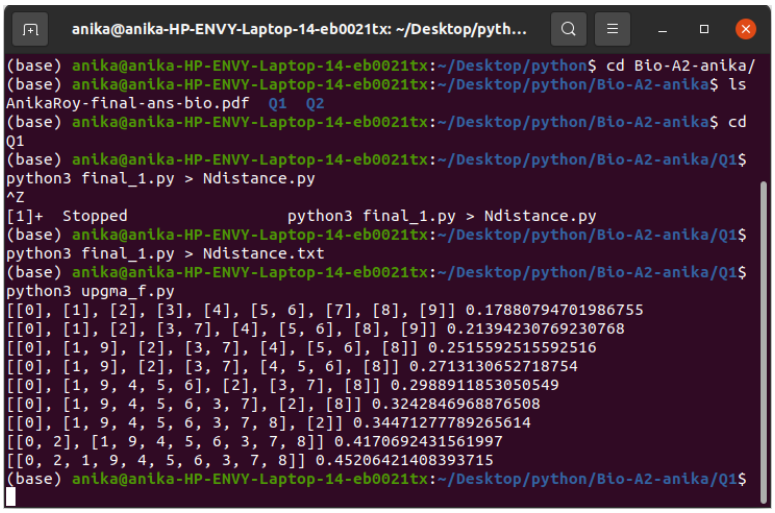
ANIKA ROY
BIOLOGY ASSIGNMENT 2 REPORT

Q1)
The Q1 folder contains final_1.py—for generating distance matrix using Nucleotide.txt.
I have assigned 0 to mismatches,insertions and deletions and 1 to matches while minimising the sequence length. I calculate the pairwise distances using dp and keep printing to the output file Ndistance.txt.
upgma_f.py reads this output and implements UPGMA algorithm giving the following output



at every stage, it shows what i have clustered together and the number at the end is the distance between the groups we just clustered which i then divide equally in the graph. I have used this output to hand draw the graph.

Q2)
The Q2 folder contains final_2.py—for generating distance matrix using Protein.txt.
Using the BLOSUM 62 table, I have assigned weights to mismatches and matches while minimising the sequence length. For all insertions and deletions, I have assigned a penalty of -1.I calculate the pairwise score using dp and keep printing the inverse(my distance parameter) to the output file Pdistance.txt.
upgma_f.py reads this output and implements UPGMA algorithm giving the following output.