

Assignment 10

1. To obtain the consensus phylogenetic tree for the given sequences, we follow the steps given in the Practical 10 ppt file and we get the following results. (The files are attached to the submission)

Files:

Seq 1:

outtree	:	seq1Outtree.pdf
nj-tree	:	work1-nj-tree.pdf
constree	:	work1-constree.pdf
nj-constree	:	work1-nj-constree.pdf

Seq 2:

outtree	:	seq2Outtree.pdf
nj-tree	:	work2-nj-tree.pdf
constree	:	work2-constree.pdf
nj-constree	:	work2-nj-constree.pdf

2. We are given some sequences, to find weight matrix for the given sequences, we need to find consensus of the sequences and then do the weight matrix using the below formula:

$$W_{ij} = \log \left(\frac{N_{ij} + Prob}{Prob \times (no. sequences + 1)} \right)$$

Where,

W_{ij} = Weight of the j^{th} element of i^{th} sequence of the weight matrix

N_{ij} = Consensus element of the j^{th} element of i^{th} sequence of the alignment matrix

Prob = Probability for that particular amino acid (1/20)

No.sequences = number of sequences

The code for getting weight matrix is:

```
import numpy as np
import pandas as pd
import math
```

```

def weight_matrix(strarr):
    AA_all=['A','C','D','E','F','G','H','I','K','L','M','N','P','Q','R','S','T','V','W','Y']
    align_mat=[[0 for i in range(len(strarr[0]))] for i in range(len(AA_all))]
    weight_mat=[[0.0 for i in range(len(strarr[0]))] for i in range(len(AA_all))]
    prob=1/20
    for i in range(len(strarr)):
        for j in range(len(strarr[i])):
            align_mat[AA_all.index(strarr[i][j])][j]+=1
    align_data=np.array(align_mat)
    print('Alignment Matrix:')
    df1=pd.DataFrame(align_data, AA_all, [i+1 for i in range(len(strarr[0]))])
    print(df1.to_string())
    for i in range(len(AA_all)):
        for j in range(len(strarr[0])):
            num=align_mat[i][j]+prob
            den=prob*(len(strarr)+1)
            fra=num/den
            weight_mat[i][j]=math.log(fra)
    weight_data=weight_mat
    print('Weight Matrix:')
    df2=pd.DataFrame(weight_data, AA_all, [i+1 for i in range(len(strarr[0]))])
    print(df2.to_string())

strings=['MVLSPADKTNVKGKVGAGHAGEYGAAAW',
        'MKRLPADPPCVKGKVKAKAGDYGATTW',
        'MALSAADKTNVKS KVG GHAGEYGAATS',
        'MVLSAADKTNVKS KAGGNAGEWWAAAW',
        'MVLSAADKTNVKS KVLNAGEFGAAAW',
        'ALLPIRTTYHKK CASGHIPEEKDLNNV',
        'DEASSLKGH HIKKLEADALLIPLSASS',]
weight_matrix(strings)

```

And the output we get is (Raw Text):

Alignment Matrix:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
25	26	27																						
A	1	1	1	0	3	5	0	0	0	0	0	0	0	1	1	1	3	1	5	0	0	0	0	5
5	3	0																						
C	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0																						
D	1	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0
0	0	0																						
E	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	5	0	0	0
0	0	0																						
F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0																						
G	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	4	2	0	0	5	0	0	4	0
0	0	0																						
H	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	1	2	0	0	0	0	0	0
0	0	0																						
I	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0
0	0	0																						
K	0	1	0	0	0	0	1	4	0	0	1	7	1	5	0	1	0	1	0	0	0	1	0	0
0	0	0																						
L	0	1	5	1	0	1	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0	1	1
0	0	0																						
M	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0																						
N	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	2	0	0	0	0	0	0
1	1	0																						
P	0	0	0	1	2	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0
0	0	0																						
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0																						

R	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O	0	0																						
S	0	0	0	5	1	0	0	0	0	0	0	0	3	0	1	0	0	0	0	0	0	0	0	1
O	1	2																						
T	0	0	0	0	0	0	1	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
l	2	0																						
V	0	3	0	0	0	0	0	0	0	0	5	0	0	0	4	0	0	0	0	0	0	0	0	0
O	0	1																						
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
O	0	4																						
Y	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
O	0	0																						

Weight Matrix:

	1	2	3	4	5	6	7	8	9		
10	11	12	13	14	15	16	17	18	19		
20	21	22	23	24	25	26	27				
A	0.965081	0.965081	0.965081	-2.079442	2.031432	2.535679	-2.079442	-2.079442	-2.079442	-	
	2.079442	-2.079442	-2.079442	-2.079442	0.965081	0.965081	0.965081	2.031432	0.965081	2.535679	-
	2.079442	-2.079442	-2.079442	-2.079442	2.535679	2.535679	2.031432	-2.079442			
C	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442		
	0.965081	-2.079442	-2.079442	0.965081	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-
	2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442			
D	0.965081	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	2.535679	-2.079442	-2.079442	-	
	2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	0.965081	-2.079442	-2.079442	-
	2.079442	0.965081	-2.079442	0.965081	-2.079442	-2.079442	-2.079442	-2.079442			
E	-2.079442	0.965081	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-	
	2.079442	-2.079442	-2.079442	-2.079442	-2.079442	0.965081	-2.079442	-2.079442	-2.079442	-2.079442	
	0.965081	2.535679	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442			
F	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-	
	2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442	-
	2.079442	-2.079442	0.965081	-2.079442	-2.079442	-2.079442	-2.079442	-2.079442			

G -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 0.965081 -2.079442 -
2.079442 -2.079442 -2.079442 1.634131 -2.079442 -2.079442 2.315008 1.634131 -2.079442 -2.079442
2.535679 -2.079442 -2.079442 2.315008 -2.079442 -2.079442 -2.079442 -2.079442
H -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 0.965081
1.634131 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 0.965081 1.634131 -2.079442 -
2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442
I -2.079442 -2.079442 -2.079442 -2.079442 0.965081 -2.079442 -2.079442 -2.079442 -2.079442 -
2.079442 0.965081 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 0.965081 -2.079442 -
2.079442 0.965081 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442
K -2.079442 0.965081 -2.079442 -2.079442 -2.079442 -2.079442 0.965081 2.315008 -2.079442 -
2.079442 0.965081 2.869318 0.965081 2.535679 -2.079442 0.965081 -2.079442 0.965081 -2.079442 -
2.079442 -2.079442 0.965081 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442
L -2.079442 0.965081 2.535679 0.965081 -2.079442 0.965081 -2.079442 -2.079442 -2.079442 -
2.079442 -2.079442 -2.079442 -2.079442 0.965081 -2.079442 0.965081 -2.079442 -2.079442 0.965081
0.965081 -2.079442 -2.079442 0.965081 0.965081 -2.079442 -2.079442 -2.079442
M 2.535679 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -
2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -
2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442
N -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442
2.315008 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 1.634131 -2.079442 -
2.079442 -2.079442 -2.079442 -2.079442 -2.079442 0.965081 0.965081 -2.079442
P -2.079442 -2.079442 -2.079442 0.965081 1.634131 -2.079442 -2.079442 0.965081 0.965081 -
2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 0.965081 -
2.079442 -2.079442 0.965081 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442
Q -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -
2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -
2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442
R -2.079442 -2.079442 0.965081 -2.079442 -2.079442 0.965081 -2.079442 -2.079442 -2.079442 -
2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -
2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442
S -2.079442 -2.079442 -2.079442 2.535679 0.965081 -2.079442 -2.079442 -2.079442 -2.079442 -
2.079442 -2.079442 -2.079442 2.031432 -2.079442 0.965081 -2.079442 -2.079442 -2.079442 -2.079442 -
2.079442 -2.079442 -2.079442 -2.079442 0.965081 -2.079442 0.965081 1.634131

```

T -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 0.965081 0.965081 2.315008 -
2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -
2.079442 -2.079442 -2.079442 -2.079442 -2.079442 0.965081 1.634131 -2.079442
V -2.079442 2.031432 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -
2.079442 2.535679 -2.079442 -2.079442 -2.079442 2.315008 -2.079442 -2.079442 -2.079442 -2.079442 -
2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 0.965081
W -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -
2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -
2.079442 -2.079442 0.965081 0.965081 -2.079442 -2.079442 -2.079442 2.315008
Y -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 0.965081 -
2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442 -
2.079442 -2.079442 2.031432 -2.079442 -2.079442 -2.079442 -2.079442 -2.079442

```

Graphical Output:

Alignment Matrix:

