Question 1

Modify equation 1 for dynamic programming to allow gaps only in sequence X.

$$M[i,j] = max \begin{pmatrix} M[i-1,j-1] + score\,(X[i], Y[j]) \\ M[i,j-1] - g \\ M[i-1,j] - g \end{pmatrix} \quad (1)$$

Question 2

The edit distance between two words is the **minimum** number of operations needed to transform one word into another. Allowed operations are:

- replacement of a single letter by another
- insertion of a single letter
- deletion of a single letter

Edit distances can be determined with dynamic programming. Write down the equation for $M[i,j]$ such that the edit distance between two sequences is given by their alignment score. (Hint: You need to define the score function and gap penalty too.)

Question 3

Perform global alignment of the protein sequences DARWIN and CRICK using equation 2:

$$M[i,j] = max \begin{pmatrix} M[i-1,j-1] + blosum62\,(X[i], Y[j]) \\ M[i,j-1] - 2 \\ M[i-1,j] - 2 \end{pmatrix} \quad (2)$$

where the value of $blosum62(X[i], Y[j])$ is the substitution score between residue $X[i]$ and $Y[j]$ according to the BLOSUM62 matrix (Figure 1). Fill in the score matrix template (including arrows) in figure 2.

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
C	9																				C
S	-1	4																			S
T	-1	1	5																		T
P	-3	-1	-1	7																	P
A	0	1	0	-1	4																A
G	-3	0	-2	-2	0	6															G
N	-3	1	0	-2	-2	0	6														N
D	-3	0	-1	-1	-2	-1	1	6													D
E	-4	0	-1	-1	-1	-2	0	2	5												E
Q	-3	0	-1	-1	-1	-2	0	0	2	5											Q
H	-3	-1	-2	-2	-2	-2	1	-1	0	0	8										H
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5									R
K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5								K
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5							M
I	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4						I
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4					L
V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4				V
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6			F
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7		Y
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11	W

Figure 1: BLOSUM62 matrix

			X[i]						
		i	0	1	2	3	4	5	6
	j		-	D	A	R	W	I	N
Y[j]	0	-							
	1	C							
	2	R							
	3	I							
	4	C							
	5	K							

Figure 2: Template for Question 3

Question 4

What is the alignment score for Question 3?

Question 5

Write down the alignment from Question 3.

Question 6

Perform local alignment between the sequences TGAGA and GAGGC, using equation 3:

$$M[i,j] = max \begin{pmatrix} M[i-1,j-1] \pm 1 \\ M[i,j-1] - 2 \\ M[i-1,j] - 2 \\ 0 \end{pmatrix} \tag{3}$$

Fill in the score matrix template (including arrows) in figure 3.

			X[i]					
		i	0	1	2	3	4	5
		j	-	T	G	A	G	A
		0	-					
		1	G					
		2	A					
Y[j]								
		3	G					
		4	G					
		5	C					

Figure 3: Template for Question 6

Question 7

Use the Waterman-Eggert method to trace back through your scoring matrix from Question 6. Write down the two highest scoring **local** alignments and their respective scores.