Chuong Vu

COMP5800

Homework 2

Prof. Byung Kim

Sequence Alignment Using Dynamic Programming

1. **Objective**

The objective of the assignment is to program a pairwise alignment using the dynamic programming approach.

1. **Project Assignment**
2. Use the scoring scheme with the following base substitution scores and *g* for gap penalty (no affine gap penality).

+---+----+----+----+----+

| | T | C | A | G |

+---+----+----+----+----+

| T | 20 | 10 | 5 | 5 |

+---+----+----+----+----+

| C | 10 | 20 | 5 | 5 |

+---+----+----+----+----+

| A | 5 | 5 | 20 | 10 |

+---+----+----+----+----+

| G | 5 | 5 | 10 | 20 |

+---+----+----+----+----+

In the core of the dynamic programming algorithm, you need to consider 5-way comparisons: indel and four base substitutions.

The gap penalty is a variable, and you can try for two different values of *g* and see if there are differences in the aligned results.

1. Use two sets of test data
   1. A T G T T A T

A T C G T A G

* 1. ATGAATGCGATTTCGGGTGGCC

TTGGCAGGACATGAAGTTCGATACGGAA

1. An example of the implementation of the algorithm:

Let the score matrix in part 2 above be denoted by *S*(*ai, bi*). For example, *S*(‘A’, ‘A’) = 20 and *S*(‘A’, ‘T’) = 5. Suppose we have two short sequences of ATG and GGAAT. We can visualized two sequences lined up vertically and horizontally in a matrix, *D*, shown below. The top row and the leftmost column are for gaps, and they are initialized to multiples of gap penalties (the gap penalty *g* of 5 is used below).

* G G A A T

- 0 -5 -10 -15 -20 -25

A -5

T -10

G -15

After the initialization, each element in the matrix (except for the top row and the leftmost column) is visited, and its score is updated to the maximum of new scores computed in three directions: vertical down, diagonal, and horizontal right, as follows:

*D*(*i,j*) = max[ *D*(*i*-1,*j*) + *g*, *D*(*i*-1, *j*-1)+ *S*(*ai, bi*), *D*(*i,j*-1)+*g*]

At the same time, we record the direction that the maximum is obtained. For example, for *D*[1,1] = max[-5-5, 0+10, -5-5] = 10 (diagonal). Going down vertically, next *D*[2,1] = max[10-5, -5+5, -10-5] = 5 (vertical). Finally in the 2nd column, *D*[3,1] = max[5-5, -10+20, -15-5] = 10 (diagonal). D matrix now looks like

* G G A A T

- 0 -5 -10 -15 -20 -25

A -5 10

T -10 5

G -15 10

For the 3rd row, *D*[1,2] = max(-10-5, -5+10, 10-5) = 5 (diagonal and right), *D*[2,2] = max(5-5, 10+5, 5-5) = 15, and *D*[3,2] = max(15-5, 5+20, 10-5) = 25. *D* matrix then looks like the following:

* G G A A T

- 0 -5 -10 -15 -20 -25

A -5 10 5

T -10 5 15

G -15 10 25

Continuing on in the same manner, the final matrix turns out to be as follows.

* G G A A T

- 0 -5 -10 -15 -20 -25

A -5 10 5 10 5 0

T -10 5 15 10 15 25

G -15 10 25 25 20 20

The final alignment is to trace back from *D*[3,5] at the lower right corner. There are two paths leading to *D*[3,5]:

G G A A T

- - A T G

and

G G A A T -

- - - A T G

1. Submit the program and a report that includes the basic core of the dynamic programming routine, and aligned results of the test data set.
2. **Works and Results**
3. Methods and Coding
   1. I’m using Needleman-Wunsch’s algorithm to write my program. The pseudo-code also mostly come from the algorithm.

The pseudo-code is compute the F (scoring) matrix and biological sequence.

for i = 0 to length(A)

F(i, 0)← d \* i

for j = 0 to length(B)

F(0, j)← d \* j

for i = 1 to length(A)

for j = 1 to length(B) {

Match← F(i - 1, j - 1) + S(Ai, Bj)

Delete← F(i - 1, j) + d

Insert← F(i, j - 1) + d

F(i, j)← max(Match, Insert, Delete)

}

For the alignment sequence, the pseudo-code is:

AlignmentA ← ""

AlignmentB ← ""

i ← length(A)

j ← length(B)

while (i > 0 or j > 0) {

if (i > 0 and j > 0 and F(i,j) == F(i-1,j-1) + S(Ai, Bj)) {

AlignmentA ← Ai + AlignmentA

AlignmentB ← Bj + AlignmentB

i ← i - 1

j ← j - 1

}

else if (i > 0 and F(i,j) == F(i-1,j) + d) {

AlignmentA ← Ai + AlignmentA

AlignmentB ← "-" + AlignmentB

i ← i - 1

} else {

AlignmentA ← "-" + AlignmentA

AlignmentB ← Bj + AlignmentB

j ← j - 1

}

}

* 1. My code in Python3 (may be work on Python 2.7). The program is based on the dynamic programming to compare the biological sequences. My programming routine in Python as below:

def CalTable(strX, strY, gap):

print("Data\_Set 1st Sequence:", strX)

print("Data\_Set 2nd Sequence:", strY)

print("Gap:", gap)

strX = "-" + strX

strY = "-" + strY

len\_strX = len(strX)

len\_strY = len(strY)

score = len\_strX\*[len\_strY\*[0]]

score\_table = []

j = 0

for row in range(0, len\_strY + 1):

score\_table.append([])

for col in range(0, len\_strX + 1):

if row == 0 and col == 0:

score\_table[0].append(" ")

elif row == 0 and col > 0:

score\_table[0].append(strX[col - 1])

else:

if col == 0:

score\_table[row].append(strY[j])

j += 1

elif row == 1 and col == 1:

score\_table[row].append(0)

elif row == 1:

score\_table[row].append(score\_table[row][col - 1] + gap)

elif col == 1:

score\_table[row].append(score\_table[row - 1][col] + gap)

else:

Match = score\_table[row - 1][col - 1] + score\_dict[(strY[row - 1], strX[col - 1])]

#print(row, col, strY[row - 1], strX[col - 1])

Insert = score\_table[row][col - 1] + gap

Delete = score\_table[row - 1][col] + gap

score\_table[row].append(max(Match, Insert, Delete))

table\_print = AsciiTable(score\_table)

table\_print.inner\_row\_border = True

# for i in range(0, len\_strX + 1):

# table\_print.justify\_columns[i] = 'center'

print(table\_print.table)

GetScore(score\_table, strX, strY, gap)

return

##########################################

def GetScore(AliTable, strX, strY, gap):

strX = '-' + strX

strY = '-' + strY

# Pseudo-code based on Needleman-Wunsch algorithm that

AlignmentA = []

AlignmentB = []

row = len(strY) - 1

col = len(strX) - 1

while(row > 1 or col > 1):

if (row > 1 and col > 1 and AliTable[row][col] == AliTable[row - 1][col - 1] + score\_dict[(strY[row], strX[col])]):

AlignmentA.insert(0, strY[row])

AlignmentB.insert(0, strX[col])

row -= 1

col -= 1

elif (row > 1 and AliTable[row][col] == AliTable[row - 1][col] + gap):

AlignmentA.insert(0, strY[row])

AlignmentB.insert(0, '-')

row -= 1

else:

AlignmentA.insert(0, '-')

AlignmentB.insert(0, strX[col])

col -= 1

print()

print("Alignment Sqeuence Based on Needleman-Wunnsch algorithm")

table\_print = AsciiTable([AlignmentA,AlignmentB])

table\_print.inner\_row\_border = True

#table\_print.justify\_columns[0] = 'center'

print(table\_print.table)

Print\_Score(AlignmentA, AlignmentB)

return

1. Results

The result will show in two data set for testing. The display below is show the scoring matrix, data set and alignment sequence.

1. First data set:

Data\_Set 1st Sequence: ATGTTAT

Data\_Set 2nd Sequence: ATCGTAG

Gap Penalty: -1

Scoring Matrix

+---+----+----+----+----+----+----+----+-----+

| | - | A | T | G | T | T | A | T |

+---+----+----+----+----+----+----+----+-----+

| - | 0 | -1 | -2 | -3 | -4 | -5 | -6 | -7 |

+---+----+----+----+----+----+----+----+-----+

| A | -1 | 20 | 19 | 18 | 17 | 16 | 15 | 14 |

+---+----+----+----+----+----+----+----+-----+

| T | -2 | 19 | 40 | 39 | 38 | 37 | 36 | 35 |

+---+----+----+----+----+----+----+----+-----+

| C | -3 | 18 | 39 | 45 | 49 | 48 | 47 | 46 |

+---+----+----+----+----+----+----+----+-----+

| G | -4 | 17 | 38 | 59 | 58 | 57 | 58 | 57 |

+---+----+----+----+----+----+----+----+-----+

| T | -5 | 16 | 37 | 58 | 79 | 78 | 77 | 78 |

+---+----+----+----+----+----+----+----+-----+

| A | -6 | 15 | 36 | 57 | 78 | 84 | 98 | 97 |

+---+----+----+----+----+----+----+----+-----+

| G | -7 | 14 | 35 | 56 | 77 | 83 | 97 | 103 |

+---+----+----+----+----+----+----+----+-----+

Alignment Sequence based on Needleman-Wunsch's algorithm

+---+---+---+---+---+---+---+---+

| A | T | C | G | - | T | A | G |

+---+---+---+---+---+---+---+---+

| A | T | - | G | T | T | A | T |

+---+---+---+---+---+---+---+---+

Matched Sequence = 5

Mismatched Sequence = 1

Gap = 2

Alignment Sequence Score = 2

------------------------------------------------------------

Data\_Set 1st Sequence: ATGTTAT

Data\_Set 2nd Sequence: ATCGTAG

Gap Penalty: -5

Scoring Matrix

+---+-----+-----+-----+-----+-----+-----+-----+-----+

| | - | A | T | G | T | T | A | T |

+---+-----+-----+-----+-----+-----+-----+-----+-----+

| - | 0 | -5 | -10 | -15 | -20 | -25 | -30 | -35 |

+---+-----+-----+-----+-----+-----+-----+-----+-----+

| A | -5 | 20 | 15 | 10 | 5 | 0 | -5 | -10 |

+---+-----+-----+-----+-----+-----+-----+-----+-----+

| T | -10 | 15 | 40 | 35 | 30 | 25 | 20 | 15 |

+---+-----+-----+-----+-----+-----+-----+-----+-----+

| C | -15 | 10 | 35 | 45 | 45 | 40 | 35 | 30 |

+---+-----+-----+-----+-----+-----+-----+-----+-----+

| G | -20 | 5 | 30 | 55 | 50 | 50 | 50 | 45 |

+---+-----+-----+-----+-----+-----+-----+-----+-----+

| T | -25 | 0 | 25 | 50 | 75 | 70 | 65 | 70 |

+---+-----+-----+-----+-----+-----+-----+-----+-----+

| A | -30 | -5 | 20 | 45 | 70 | 80 | 90 | 85 |

+---+-----+-----+-----+-----+-----+-----+-----+-----+

| G | -35 | -10 | 15 | 40 | 65 | 75 | 90 | 95 |

+---+-----+-----+-----+-----+-----+-----+-----+-----+

Alignment Sequence based on Needleman-Wunsch's algorithm

+---+---+---+---+---+---+---+

| A | T | C | G | T | A | G |

+---+---+---+---+---+---+---+

| A | T | G | T | T | A | T |

+---+---+---+---+---+---+---+

Matched Sequence = 4

Mismatched Sequence = 3

Gap = 0

Alignment Sequence Score = 1

1. Second Data Set

Data\_Set 1st Sequence: ATGAATGCGATTTCGGGTGGCC

Data\_Set 2nd Sequence: TTGGCAGGACATGAAGTTCGATACGGAA

Gap Penalty: -1

Scoring Matrix

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| | - | A | T | G | A | A | T | G | C | G | A | T | T | T | C | G | G | G | T | G | G | C | C |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| - | 0 | -1 | -2 | -3 | -4 | -5 | -6 | -7 | -8 | -9 | -10 | -11 | -12 | -13 | -14 | -15 | -16 | -17 | -18 | -19 | -20 | -21 | -22 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| T | -1 | 5 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | -1 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| T | -2 | 4 | 25 | 24 | 23 | 22 | 36 | 35 | 34 | 33 | 32 | 31 | 30 | 29 | 28 | 27 | 26 | 25 | 24 | 23 | 22 | 21 | 20 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| G | -3 | 8 | 24 | 45 | 44 | 43 | 42 | 56 | 55 | 54 | 53 | 52 | 51 | 50 | 49 | 48 | 47 | 46 | 45 | 44 | 43 | 42 | 41 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| G | -4 | 7 | 23 | 44 | 55 | 54 | 53 | 62 | 61 | 75 | 74 | 73 | 72 | 71 | 70 | 69 | 68 | 67 | 66 | 65 | 64 | 63 | 62 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| C | -5 | 6 | 22 | 43 | 54 | 60 | 64 | 63 | 82 | 81 | 80 | 84 | 83 | 82 | 91 | 90 | 89 | 88 | 87 | 86 | 85 | 84 | 83 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| A | -6 | 15 | 21 | 42 | 63 | 74 | 73 | 74 | 81 | 92 | 101 | 100 | 99 | 98 | 97 | 101 | 100 | 99 | 98 | 97 | 96 | 95 | 94 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| G | -7 | 14 | 20 | 41 | 62 | 73 | 79 | 93 | 92 | 101 | 102 | 106 | 105 | 104 | 103 | 117 | 121 | 120 | 119 | 118 | 117 | 116 | 115 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| G | -8 | 13 | 19 | 40 | 61 | 72 | 78 | 99 | 98 | 112 | 111 | 110 | 111 | 110 | 109 | 123 | 137 | 141 | 140 | 139 | 138 | 137 | 136 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| A | -9 | 12 | 18 | 39 | 60 | 81 | 80 | 98 | 104 | 111 | 132 | 131 | 130 | 129 | 128 | 127 | 136 | 147 | 146 | 150 | 149 | 148 | 147 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| C | -10 | 11 | 22 | 38 | 59 | 80 | 91 | 97 | 118 | 117 | 131 | 142 | 141 | 140 | 149 | 148 | 147 | 146 | 157 | 156 | 155 | 169 | 168 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| A | -11 | 10 | 21 | 37 | 58 | 79 | 90 | 101 | 117 | 128 | 137 | 141 | 147 | 146 | 148 | 159 | 158 | 157 | 156 | 167 | 166 | 168 | 174 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| T | -12 | 9 | 30 | 36 | 57 | 78 | 99 | 100 | 116 | 127 | 136 | 157 | 161 | 167 | 166 | 165 | 164 | 163 | 177 | 176 | 175 | 176 | 178 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| G | -13 | 8 | 29 | 50 | 56 | 77 | 98 | 119 | 118 | 136 | 137 | 156 | 162 | 166 | 172 | 186 | 185 | 184 | 183 | 197 | 196 | 195 | 194 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| A | -14 | 7 | 28 | 49 | 70 | 76 | 97 | 118 | 124 | 135 | 156 | 155 | 161 | 167 | 171 | 185 | 196 | 195 | 194 | 196 | 207 | 206 | 205 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| A | -15 | 6 | 27 | 48 | 69 | 90 | 96 | 117 | 123 | 134 | 155 | 161 | 160 | 166 | 172 | 184 | 195 | 206 | 205 | 204 | 206 | 212 | 211 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| G | -16 | 5 | 26 | 47 | 68 | 89 | 95 | 116 | 122 | 143 | 154 | 160 | 166 | 165 | 171 | 192 | 204 | 215 | 214 | 225 | 224 | 223 | 222 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| T | -17 | 4 | 25 | 46 | 67 | 88 | 109 | 115 | 126 | 142 | 153 | 174 | 180 | 186 | 185 | 191 | 203 | 214 | 235 | 234 | 233 | 234 | 233 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| T | -18 | 3 | 24 | 45 | 66 | 87 | 108 | 114 | 125 | 141 | 152 | 173 | 194 | 200 | 199 | 198 | 202 | 213 | 234 | 240 | 239 | 243 | 244 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| C | -19 | 2 | 23 | 44 | 65 | 86 | 107 | 113 | 134 | 140 | 151 | 172 | 193 | 204 | 220 | 219 | 218 | 217 | 233 | 239 | 245 | 259 | 263 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| G | -20 | 1 | 22 | 43 | 64 | 85 | 106 | 127 | 133 | 154 | 153 | 171 | 192 | 203 | 219 | 240 | 239 | 238 | 237 | 253 | 259 | 258 | 264 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| A | -21 | 0 | 21 | 42 | 63 | 84 | 105 | 126 | 132 | 153 | 174 | 173 | 191 | 202 | 218 | 239 | 250 | 249 | 248 | 252 | 263 | 264 | 263 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| T | -22 | -1 | 20 | 41 | 62 | 83 | 104 | 125 | 136 | 152 | 173 | 194 | 193 | 211 | 217 | 238 | 249 | 255 | 269 | 268 | 267 | 273 | 274 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| A | -23 | -2 | 19 | 40 | 61 | 82 | 103 | 124 | 135 | 151 | 172 | 193 | 199 | 210 | 216 | 237 | 248 | 259 | 268 | 279 | 278 | 277 | 278 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| C | -24 | -3 | 18 | 39 | 60 | 81 | 102 | 123 | 144 | 150 | 171 | 192 | 203 | 209 | 230 | 236 | 247 | 258 | 269 | 278 | 284 | 298 | 297 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| G | -25 | -4 | 17 | 38 | 59 | 80 | 101 | 122 | 143 | 164 | 170 | 191 | 202 | 208 | 229 | 250 | 256 | 267 | 268 | 289 | 298 | 297 | 303 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| G | -26 | -5 | 16 | 37 | 58 | 79 | 100 | 121 | 142 | 163 | 174 | 190 | 201 | 207 | 228 | 249 | 270 | 276 | 275 | 288 | 309 | 308 | 307 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| A | -27 | -6 | 15 | 36 | 57 | 78 | 99 | 120 | 141 | 162 | 183 | 189 | 200 | 206 | 227 | 248 | 269 | 280 | 281 | 287 | 308 | 314 | 313 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

| A | -28 | -7 | 14 | 35 | 56 | 77 | 98 | 119 | 140 | 161 | 182 | 188 | 199 | 205 | 226 | 247 | 268 | 279 | 285 | 291 | 307 | 313 | 319 |

+---+-----+----+----+----+----+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

Alignment Sequence based on Needleman-Wunsch's algorithm

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

| T | T | G | G | C | A | G | G | A | C | A | T | G | A | A | G | T | T | C | G | A | T | A | C | G | G | A | A |

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

| A | T | G | A | - | A | T | G | - | C | - | - | G | - | A | T | T | T | C | G | G | - | G | T | G | G | C | C |

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

Matched Sequence = 13

Mismatched Sequence = 9

Gap = 6

Alignment Sequence Score = -2

------------------------------------------------------------

Data\_Set 1st Sequence: ATGAATGCGATTTCGGGTGGCC

Data\_Set 2nd Sequence: TTGGCAGGACATGAAGTTCGATACGGAA

Gap Penalty: -5

Scoring Matrix

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| | - | A | T | G | A | A | T | G | C | G | A | T | T | T | C | G | G | G | T | G | G | C | C |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| - | 0 | -5 | -10 | -15 | -20 | -25 | -30 | -35 | -40 | -45 | -50 | -55 | -60 | -65 | -70 | -75 | -80 | -85 | -90 | -95 | -100 | -105 | -110 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| T | -5 | 5 | 15 | 10 | 5 | 0 | -5 | -10 | -15 | -20 | -25 | -30 | -35 | -40 | -45 | -50 | -55 | -60 | -65 | -70 | -75 | -80 | -85 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| T | -10 | 0 | 25 | 20 | 15 | 10 | 20 | 15 | 10 | 5 | 0 | -5 | -10 | -15 | -20 | -25 | -30 | -35 | -40 | -45 | -50 | -55 | -60 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| G | -15 | 0 | 20 | 45 | 40 | 35 | 30 | 40 | 35 | 30 | 25 | 20 | 15 | 10 | 5 | 0 | -5 | -10 | -15 | -20 | -25 | -30 | -35 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| G | -20 | -5 | 15 | 40 | 55 | 50 | 45 | 50 | 45 | 55 | 50 | 45 | 40 | 35 | 30 | 25 | 20 | 15 | 10 | 5 | 0 | -5 | -10 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| C | -25 | -10 | 10 | 35 | 50 | 60 | 60 | 55 | 70 | 65 | 60 | 60 | 55 | 50 | 55 | 50 | 45 | 40 | 35 | 30 | 25 | 20 | 15 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| A | -30 | -5 | 5 | 30 | 55 | 70 | 65 | 70 | 65 | 80 | 85 | 80 | 75 | 70 | 65 | 65 | 60 | 55 | 50 | 45 | 40 | 35 | 30 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| G | -35 | -10 | 0 | 25 | 50 | 65 | 75 | 85 | 80 | 85 | 90 | 90 | 85 | 80 | 75 | 85 | 85 | 80 | 75 | 70 | 65 | 60 | 55 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| G | -40 | -15 | -5 | 20 | 45 | 60 | 70 | 95 | 90 | 100 | 95 | 95 | 95 | 90 | 85 | 95 | 105 | 105 | 100 | 95 | 90 | 85 | 80 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| A | -45 | -20 | -10 | 15 | 40 | 65 | 65 | 90 | 100 | 100 | 120 | 115 | 110 | 105 | 100 | 95 | 105 | 115 | 110 | 110 | 105 | 100 | 95 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| C | -50 | -25 | -10 | 10 | 35 | 60 | 75 | 85 | 110 | 105 | 115 | 130 | 125 | 120 | 125 | 120 | 115 | 110 | 125 | 120 | 115 | 125 | 120 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| A | -55 | -30 | -15 | 5 | 30 | 55 | 70 | 85 | 105 | 120 | 125 | 125 | 135 | 130 | 125 | 135 | 130 | 125 | 120 | 135 | 130 | 125 | 130 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| T | -60 | -35 | -10 | 0 | 25 | 50 | 75 | 80 | 100 | 115 | 125 | 145 | 145 | 155 | 150 | 145 | 140 | 135 | 145 | 140 | 140 | 140 | 135 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| G | -65 | -40 | -15 | 10 | 20 | 45 | 70 | 95 | 95 | 120 | 125 | 140 | 150 | 150 | 160 | 170 | 165 | 160 | 155 | 165 | 160 | 155 | 150 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| A | -70 | -45 | -20 | 5 | 30 | 40 | 65 | 90 | 100 | 115 | 140 | 135 | 145 | 155 | 155 | 170 | 180 | 175 | 170 | 165 | 175 | 170 | 165 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| A | -75 | -50 | -25 | 0 | 25 | 50 | 60 | 85 | 95 | 110 | 135 | 145 | 140 | 150 | 160 | 165 | 180 | 190 | 185 | 180 | 175 | 180 | 175 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| G | -80 | -55 | -30 | -5 | 20 | 45 | 55 | 80 | 90 | 115 | 130 | 140 | 150 | 145 | 155 | 180 | 185 | 200 | 195 | 205 | 200 | 195 | 190 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| T | -85 | -60 | -35 | -10 | 15 | 40 | 65 | 75 | 90 | 110 | 125 | 150 | 160 | 170 | 165 | 175 | 185 | 195 | 220 | 215 | 210 | 210 | 205 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| T | -90 | -65 | -40 | -15 | 10 | 35 | 60 | 70 | 85 | 105 | 120 | 145 | 170 | 180 | 180 | 175 | 180 | 190 | 215 | 225 | 220 | 220 | 220 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| C | -95 | -70 | -45 | -20 | 5 | 30 | 55 | 65 | 90 | 100 | 115 | 140 | 165 | 180 | 200 | 195 | 190 | 185 | 210 | 220 | 230 | 240 | 240 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| G | -100 | -75 | -50 | -25 | 0 | 25 | 50 | 75 | 85 | 110 | 110 | 135 | 160 | 175 | 195 | 220 | 215 | 210 | 205 | 230 | 240 | 235 | 245 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| A | -105 | -80 | -55 | -30 | -5 | 20 | 45 | 70 | 80 | 105 | 130 | 130 | 155 | 170 | 190 | 215 | 230 | 225 | 220 | 225 | 240 | 245 | 240 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| T | -110 | -85 | -60 | -35 | -10 | 15 | 40 | 65 | 80 | 100 | 125 | 150 | 150 | 175 | 185 | 210 | 225 | 235 | 245 | 240 | 235 | 250 | 255 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| A | -115 | -90 | -65 | -40 | -15 | 10 | 35 | 60 | 75 | 95 | 120 | 145 | 155 | 170 | 180 | 205 | 220 | 235 | 240 | 255 | 250 | 245 | 255 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| C | -120 | -95 | -70 | -45 | -20 | 5 | 30 | 55 | 80 | 90 | 115 | 140 | 155 | 165 | 190 | 200 | 215 | 230 | 245 | 250 | 260 | 270 | 265 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| G | -125 | -100 | -75 | -50 | -25 | 0 | 25 | 50 | 75 | 100 | 110 | 135 | 150 | 160 | 185 | 210 | 220 | 235 | 240 | 265 | 270 | 265 | 275 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| G | -130 | -105 | -80 | -55 | -30 | -5 | 20 | 45 | 70 | 95 | 110 | 130 | 145 | 155 | 180 | 205 | 230 | 240 | 240 | 260 | 285 | 280 | 275 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| A | -135 | -110 | -85 | -60 | -35 | -10 | 15 | 40 | 65 | 90 | 115 | 125 | 140 | 150 | 175 | 200 | 225 | 240 | 245 | 255 | 280 | 290 | 285 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

| A | -140 | -115 | -90 | -65 | -40 | -15 | 10 | 35 | 60 | 85 | 110 | 120 | 135 | 145 | 170 | 195 | 220 | 235 | 245 | 255 | 275 | 285 | 295 |

+---+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+------+------+------+

Alignment Sequence based on Needleman-Wunsch's algorithm

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

| T | T | G | G | C | A | G | G | A | C | A | T | G | A | A | G | T | T | C | G | A | T | A | C | G | G | A | A |

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

| A | T | G | A | - | A | T | G | - | C | - | - | G | - | A | T | T | T | C | G | G | - | G | T | G | G | C | C |

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

Matched Sequence = 13

Mismatched Sequence = 9

Gap = 6

Alignment Sequence Score = -2

1. **Discussion**

By following the Needle-Wunsch’s algorithm, I manually created the scoring matrix based on the given data sets. At first I filled out the scoring matrix table based on the scoring scheme and gap penalty. It filled from top left to bottom right, and then I traverse from bottom right to top left to get the Alignment Sequence. First alignment is checks the diagonal first, up second and left for pick the sequence cell. Second, alignment checks the up first, diagonal second and life last. Both alignment sequences came up with the best score it can get.