

INF281 Exercise 03

1. Affine gap penalty

Affine gap penalties are often preferable ways to calculate gap scores than linear penalties. A gap with length l can be calculated as: $g_l = g_{open} + (l - 1) * g_{extend}$.

Use the following scoring scheme and gap penalties to answer the questions.

Scoring scheme:

$$R_{ab} = 1 \text{ for } a = b$$

$$R_{ab} = 0 \text{ for } a \neq b$$

$$g_{open} = 1, g_{extend} = 0.1$$

- (a) What is the gap penalty when $l = 2$.

Solution: $g_{l=2} = 1 + (2 - 1) * 0.1 = 1.1$

- (b) Calculate the scores of the alignments.

1. q: CAGCT
d: CT--T

Solution: 0.9

2. q: CAGCT
d: C-T-T

Solution: 0

3. q: CCT--
d: ---CT

Solution: -2.3

2. Affine gap with single DP table

You need to check extra cells in addition to the adjacent cells of H when finding an optimal alignment with affine gap penalties.

Scoring scheme:

$$R_{ab} = 1 \text{ for } a = b$$

$$R_{ab} = 0 \text{ for } a \neq b$$

$$g_{open} = 1, g_{extend} = 0.1$$

		C	G
		0	-1
C		-1	1
A		-1.1	0

Assume we want to update $H_{2,2}$ and answer the following questions.

- (a) Calculate $H_{1,1} + R_{q_2,d_2}$.

Solution: $1 + 0 = 1$

- (b) Calculate $\max_{1 \leq l \leq 2} (H_{2,2-l} - g_l)$.

Solution: $\max(H_{2,1} - g_{l=1}, H_{2,0} - g_{l=2}) = \max(01, -1.11.1) = -1$

- (c) Calculate $\max_{1 \leq l \leq 2} (H_{2-l,2} - g_l)$.

Solution: $\max(H_{1,2} - g_{l=1}, H_{0,2} - g_{l=2}) = \max(01, -1.11.1) = -1$

- (d) What is the score of $H_{2,2}$.

Solution: $\max(1, 1, -1) = 1$

3. Initialization for affine gap penalty

Initialize the following tables when $g_{open} = 10$ and $g_{extend} = 1$.

E

	T	G	C	
	0	-10	-11	-12
A	-10			
A	-11			

F

	T	G	C	
	0	-10	-11	-12
A	-11			
A	-12			

G

	T	G	C	
	0	-10	-11	-12
A	-10			
A	-11			

4. Affine gap with three DP tables

Use the following scoring scheme and gap penalties to find the optimal alignment score of two sequences $q = AG$ and $d = GGGC$.

Scoring scheme:

$$R_{ab} = 1 \text{ for } a = b$$

$$R_{ab} = 0 \text{ for } a \neq b$$

$$g_{open} = 1, g_{extend} = 0.1$$

(a) Fill all blank cells in the DP tables E, F, and G.

		E							F				
			G	G	G	C				G	G	G	C
A		0	-1	-1.1	-1.2	-1.3	A		0	-1	-1.1	-1.2	-1.3
	-1	-2	-2.1	-2.2	-2.3			-1	-2	-1	-1.1	-1.2	
G	-1.1	-1	-2	-2.1	-2.2		G	-1.1	-2.1	-1	0	-0.1	

		G				
			G	G	G	C
A	-1	-1	0	-1	-1.1	-1.2
G	-1.1	0	1	0	-1.1	

(b) What is the optimal score?

Solution: -0.1

5. Backtrack with affine gap penalty

Perform backtracking on E, F, and G tables to find the optimal alignment. The cells with double border should be visited during backtracking.

		E							F				
			A	C	G	T				A	C	G	T
C							C						
G													

		G				
			A	C	G	T
C						
G						

(a) Write the optimal alignment.

Solution:

q: -CG-

d: ACGT

6. Sequence distance with DP

DP can be used to calculate the edit distance (Levenshtein distance) between two sequences.

Scoring scheme:

$$R_{ab} = 0 \text{ for } a = b$$

$$R_{ab} = -1 \text{ for } a \neq b$$

$$g = 1$$

With the scoring scheme above, the edit distance d is calculated as $1 * T$ where T is the optimal score of the DP.

Find the edit distance between two sequences $q = AG$ and $d = ACG$.

(a) Fill the DP table.

$q \backslash d$		A	C	T
	0	-1	-2	-3
A	-1	0	-1	-2
G	-2	-1	-1	-2

(b) What is the edit distance between q and d ?

Solution:

2