

Needleman-Wunsch Pairwise Alignment

Worked out example

- $S = ACA$
- $T = CAT$
- Assume substitutions and indels have unit cost
- Warm-up:
 - If the alignment were ungapped (given above), what would the cost be?
 - Can you find a less costly alignment?

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

	-	C	A	T
-				
A				
C				
A				

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

	-	C	A	T
-				
A				
C				
A				

What about $M(0,j)$ and $M(i,0)$? (Boundary cases?)

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

	-	C	A	T
-	0	1	2	3
A	1			
C	2			
A	3			

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

What is $M(1,1)$?

	-	C	A	T
-	0	1	2	3
A	1			
C	2			
A	3			

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

What is $M(1,1)$?

	-	C	A	T
-	0	1	2	3
A	1	1		
C	2			
A	3			

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

What is $M(1,2)$?

	-	C	A	T
-	0	1	2	3
A	1	1	1	
C	2			
A	3			

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

What is $M(1,3)$?

	-	C	A	T
-	0	1	2	3
A	1	1	1	2
C	2			
A	3			

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

Filling in the rest

	-	C	A	T
-	0	1	2	3
A	1	1	1	2
C	2	1		
A	3			

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

Filling in the rest

	-	C	A	T
-	0	1	2	3
A	1	1	1	2
C	2	1	2	
A	3			

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

Filling in the rest

	-	C	A	T
-	0	1	2	3
A	1	1	1	2
C	2	1	2	2
A	3			

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

Filling in the rest

	-	C	A	T
-	0	1	2	3
A	1	1	1	2
C	2	1	2	2
A	3	2		

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

Filling in the rest

	-	C	A	T
-	0	1	2	3
A	1	1	1	2
C	2	1	2	2
A	3	2	1	

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

Filling in the rest

	-	C	A	T
-	0	1	2	3
A	1	1	1	2
C	2	1	2	2
A	3	2	1	2

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

	-	C	A	T
-	0	1	2	3
A	1	1	1	2
C	2	1	2	2
A	3	2	1	2

What is the edit distance? – see bottom right corner.

What is the pairwise alignment yielding that distance? - backtracing

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

	-	C	A	T
-	0	1	2	3
A	1	1	1	2
C	2	1	2	2
A	3	2	1	2

What is the edit distance? – see bottom right corner.

What is the pairwise alignment yielding that distance? - backtracing

Worked out example

- $S = ACA$
- $T = CAT$

Remember: the (i,j) entry is the cost of aligning the prefix S_i to prefix T_j and therefore $M(i,j)$ for $i,j > 0$ is the minimum of the following three values:

$$M(i-1,j)+1,$$

$$M(i,j-1)+1,$$

$$M(i-1,j-1)+ \text{Mismatch}(s_i,t_j)$$

where we write $S = s_1s_2s_3$ and $T = t_1t_2t_3$

	-	C	A	T
-	0	1	2	3
A	1	1	1	2
C	2	1	2	2
A	3	2	1	2

What is the edit distance? – see bottom right corner.

What is the pairwise alignment yielding that distance? - backtracing

Pairwise alignment: $A C A -$
 $- C A T$