

INF281 Exercise 02 solutions

1. DP table cell update rules

Dynamic programming (DP) is an algorithm that uses table cells to memorize the sub-solutions of the target solution. DP requires three candidate scores and selects the maximum score among them when updating a cell.

$$H_{i,j}^{(0)} = H_{i-1,j} - g \quad (\text{vertical})$$

$$H_{i,j}^{(1)} = H_{i,j-1} - g \quad (\text{horizontal})$$

$$H_{i,j}^{(2)} = H_{i-1,j-1} + R_{a,b} \quad (\text{diagonal})$$

Use the simple scoring scheme below to calculate $H_{i,j}$ in Table A and B.

Scoring scheme:

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

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

$$g = 1$$

(a) Table A

| | | | |
|---|--|----|-----------|
| | | | C |
| | | 0 | -1 |
| C | | -1 | $H_{i,j}$ |

Solution: 1

(b) Table B

| | | | |
|---|--|----|-----------|
| | | C | A |
| | | 0 | 2 |
| A | | | |
| G | | -1 | $H_{i,j}$ |

Solution: 1

2. DP initialization

Initialization is the first step of the DP procedures.

- (a) Initialize the DP table with gap penalty 3.

| q\d | | C | A | T |
|-----|----|----|----|----|
| | 0 | -3 | -6 | -9 |
| C | -3 | | | |
| A | -6 | | | |

3. DP global alignment

The score of optimal global alignment is found in the cell of the bottom-right corner after updating all cells.

Scoring scheme:

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

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

$$g = 1$$

- (a) Use the simple scoring scheme and fill the empty cells with appropriate scores.

| q\d | | C | A | T |
|-----|----|----|----|----|
| | 0 | -1 | -2 | -3 |
| C | -1 | 1 | 0 | -1 |
| A | -2 | 0 | 2 | 1 |
| G | -3 | -1 | 1 | 2 |
| C | -4 | -2 | 0 | 1 |
| T | -5 | -3 | -1 | 1 |

- (b) What is the optimal score of the alignment?

Solution: 1

4. DP backtrack

Backtracking is a process to find the alignment with the optical score. It requires re-calculations of the three candidate scores.

Scoring scheme:

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

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

$$g = 1$$

- (a) Which type of candidate score – vertical, horizontal, or diagonal – is used to update the cell with a double border? Assume that the simple scoring scheme has been used.

- Table 1

| | C | A |
|---|----|---|
| A | 0 | 2 |
| G | -1 | 1 |

Solution: Vertical

- Table 2

| | A | T |
|---|----|---|
| C | 0 | 1 |
| T | -1 | 1 |

Solution: Diagonal

- (b) Use backtracking to find the optimal global alignment.

| q\d | | C | A | T |
|-----|----|----|----|----|
| | 0 | -1 | -2 | -3 |
| C | -1 | 1 | 0 | -1 |
| A | -2 | 0 | 2 | 1 |

Solution:

q: CA-

d: CAT

5. DP with score matrix

Use the score matrix below with gap penalty $g = 1$ and answer the following questions.

| | C | G | A | T |
|---|---|---|---|---|
| C | 1 | 0 | 0 | 0 |
| G | | 1 | 1 | 0 |
| A | | | 1 | 0 |
| T | | | | 1 |

(a) Calculate the alignment score.

- Alignment 1

q: ATGCT

d: CA--T

Solution: 1

- Alignment 2

q: CAGCT

d: C-A-T

Solution: 1

(b) Calculate the score of $H_{i,j}$.

- Table A

| | | | |
|---|--|----|-----------|
| | | | C |
| | | 0 | -1 |
| C | | -1 | $H_{i,j}$ |

Solution: 1

- Table B

| | | | |
|---|--|----|-----------|
| | | C | A |
| A | | 0 | 2 |
| G | | -1 | $H_{i,j}$ |

Solution: 1

- (c) Fill the empty cells with appropriate scores in the DP table. What is the optimal alignment score?

| q\d | | C | A | T |
|-----|----|----------|----------|----------|
| | 0 | -1 | -2 | -3 |
| C | -1 | 1 | 0 | -1 |
| A | -2 | 0 | 2 | 1 |
| G | -3 | -1 | 1 | 2 |
| C | -4 | -2 | 0 | 1 |
| T | -5 | -3 | -1 | 1 |

Solution: 1

- (d) There are two different alignments that give the same optimal score in the solution above. Specify both of them.

Solution:

q: CAGCT

d: CA--T

q: CAGCT

d: C-A-T