

Tarea 2

Genómica computacional 2023-I

1. Alineamientos de secuencias

1. Con el siguiente esquema de punaje $\sigma = 2$, $\delta(a, b) = -1$ y $\delta(a, a) = 1$ haz los siguientes alineamiento

a) g1=CTTAGA y g2=GTAA

b) g1 = ACCCTACCT y g2 = AGCCTCT

2. Implementa el algoritmo Needleman-Wunsch como lo vimos en clase para computar el alineamiento global entre dos secuencias. Tu implementación deberá reportar el *score* máximo en la entrada $S_{n,m}$ así como las secuencias apareadas. Una consideración que deberás tomar en cuenta para la reconstrucción es que en caso de que en una entrada de la matriz tengas más de una opción para alinear (i.e. arriba y en diagonal o izquierda y en diagonal), solo considerarás la opción diagonal.

Prueba tu implementación con las secuencias del primer ejercicio.

3. Realiza un alineamiento usando Smith-Waterman de las secuencias g1=AGCGTAG, y g2=CTCGTG.

2. Información mutua

1. Recordando que la definición de **entropía de Shannon** de la variable aleatoria $X = x_1, x_2, x_3, \dots, x_n$ es como sigue

$$H(X) = -\sum p(x_i) \cdot \log_2(p(x_i))$$

responde las siguientes preguntas.

- a) ¿Cuál es la entropía de Shannon en bits de las siguientes cadenas? En cada caso reporta la distribución de los símbolos que las conforman

1) 001011110101010010000000101011

2) 100111100111101110111011100111

3) 000000000000000111111111111111

- b) Considerando que la función de Información Mutua se puede calcular como

$$I(X : Y) = H(X) + H(Y) - H(X, Y)$$

donde $H(X, Y)$ es la entropía conjunta, calcula la información mutua para las siguientes secuencias e indica en una tabla de contingencias entre las variables la correlación entre cada una de ellas.

- 1) 00101111010101111000000010101101010111111010000101
 - 2) 00110111011101111000011110101001010111111010000101
 - 3) 10010110011110111011101110011010111010100010111010
 - 4) 10010011101110111011101110010100101010100010111010
 - 5) 10010111101010111011000001010100010101000010111010
- c) Reproduce la gráfica de un proceso Bernoulli. Para esta pregunta describe la construcción matemática o el código usado para tal reproducción.
¿Cuál es su interpretación?