

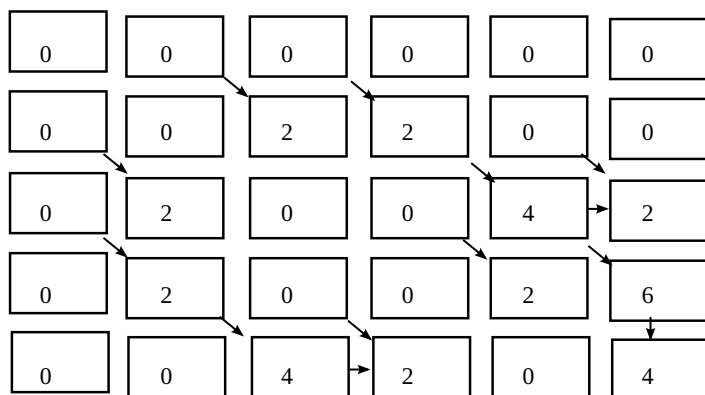
Metody Bioinformatyki, część informatyczna, 6 lutego 2021 czas: 50 min

- Proszę nie edytować tego pliku. Tutaj są tylko treści zadań.
- Odpowiedzi proszę umieszczać w pliku o nazwie 'mbi-nazwisko.txt'. Proszę pobrać ten plik, zmienić mu nazwę wpisując własne nazwisko bez użycia polskich znaków. Dodatkowo proszę wpisać swoje imię i nazwisko w pierwszej linii pliku.
- Na koniec proszę pobrać i wypełnić oświadczenie (plik 'oświadczenie-nazwisko.doc').
- **Proszę pamiętać, że przesłanie pliku na serwer należy zatwierdzić (przycisk 'Zatwierdź', lub 'Prześlij')**
- Egzamin składa się z 5 zadań. Treść każdego zadania jest na oddzielnej stronie. Pytania są w pliku z pytaniami (mbi-nazwisko.txt).

Zadanie 1 (8 pkt)

Poniżej przedstawiono macierz dla algorytmu, który dostarcza uliniowanie lokalne (algorytm Smitha-Watermana) dwóch sekwencji. Macierz podobieństwa pokazano obok. Stosujemy liniową karę za przerwę.

	<b>A</b>	<b>B</b>	<b>C</b>
<b>A</b>	2	-2	-2
<b>B</b>	-2	2	-2
<b>C</b>	-2	-2	2



**Zadanie 2 (8 pkt)**

Sekwencję AAAAABBABABAA użyto do budowy grafu de Bruijna (bez wag w krawędziach) 4 rzędu (w wierzchołkach są sekwencje o długości 3 symboli) oraz 5 rzędu (w wierzchołkach są sekwencje o długości 4 symboli).

**Zadanie 3 (8 pkt)**

Opracowano test X, który zwraca prawdopodobieństwo wystąpienia choroby. Wyniki testu są przedstawione w tabeli. Narysuj krzywą ROC dla tego testu.

osoba	stan	wynik testu
A	zdrowa	0.1
B	zdrowa	0.2
C	zdrowa	0.3
D	chora	0.4
E	chora	0.5
F	zdrowa	0.6
G	zdrowa	0.7
H	chora	0.8
I	chora	0.9
J	chora	0.9

### Zadanie 4 (8 pkt)

Mając fragment sekwencji kodującej genu *ACTG2* w genomie referencyjnym:

```
>hg38_dna range=chr2:73901312-73901338 5'pad=0 3'pad=0 strand=+
ATGTTGTAAGAGGAGACCACCGCTC
```

oraz sekwencję pacjenta, o której wiemy, że mapuje się w tym samym regionie, tj. rozpoczyna się na pozycji chr2:73901312

```
> patient1
ATGGGTGAATAGGAGACCACCGCACTC
```

oraz zakładając pierwszą ramkę odczytu (pierwszy kodon odpowiada pierwszemu aminokwasowi) wykonaj polecenia opisane w pliku z odpowiedziami.

Poniżej tablica kodonów.

		Second base							
		U	C	A	G				
First base	U	UUU } Phenylalanine <b>F</b> UUC } UUA } Leucine <b>L</b> UUG }	UCU } UCC } Serine <b>S</b> UCA } UCG }	UAU } Tyrosine <b>Y</b> UAC } UAA } Stop codon UAG } Stop codon	UGU } Cysteine <b>C</b> UGC } UGA } Stop codon UGG } Tryptophan <b>W</b>	U	C	A	G
	C	CUU } CUC } Leucine <b>L</b> CUA } CUG }	CCU } CCC } Proline <b>P</b> CCA } CCG }	CAU } Histidine <b>H</b> CAC } CAA } Glutamine <b>Q</b> CAG }	CGU } CGC } Arginine <b>R</b> CGA } CGG }	U	C	A	G
	A	AUU } Isoleucine <b>I</b> AUC } AUA } AUG } Methionine start codon <b>M</b>	ACU } ACC } Threonine <b>T</b> ACA } ACG }	AAU } Asparagine <b>N</b> AAC } AAA } Lysine <b>K</b> AAG }	AGU } Serine <b>S</b> AGC } AGA } Arginine <b>R</b> AGG }	U	C	A	G
	G	GUU } Valine <b>V</b> GUC } GUA } GUG }	GCU } GCC } Alanine <b>A</b> GCA } GCG }	GAU } Aspartic acid <b>D</b> GAC } GAA } Glutamic acid <b>E</b> GAG }	GGU } GGC } Glycine <b>G</b> GGA } GGG }	U	C	A	G
						Third base			

**Zadanie 5 (8 pkt)**

Posługujemy się monetami A, B i C, obserwując sekwencje rzutów (orły i reszki). Zakładając, że to doświadczenie jest opisywane ukrytym modelem Markowa przedstawionym niżej, podaj sekwencję monet, która najprawdopodobniej była użyta, gdy obserwujemy sekwencję rzutów *OOOR*.

$$Q = \{A, B, C\}$$
$$V = \{O, R\}$$
$$P_A = 1, P_B = 0, P_C = 0$$

	A	B	C
A	$\frac{3}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
B	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{1}{5}$
C	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{3}{5}$

	O	R
A	$\frac{1}{2}$	$\frac{1}{2}$
B	1	0
C	0	1