

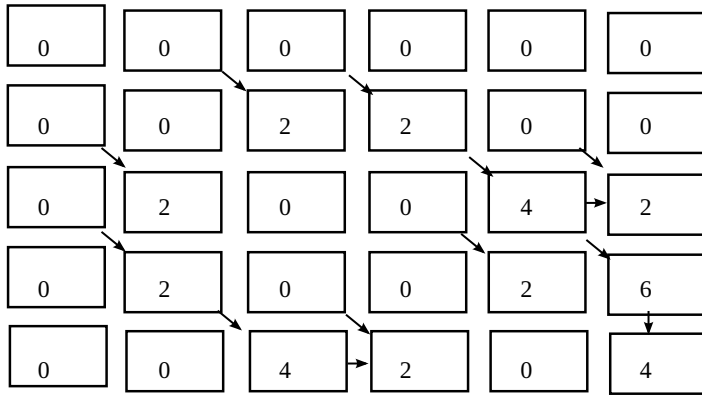
Metody Bioinformatyki, część informatyczna, 6 lutego 2021 czas: 50 min

- 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)

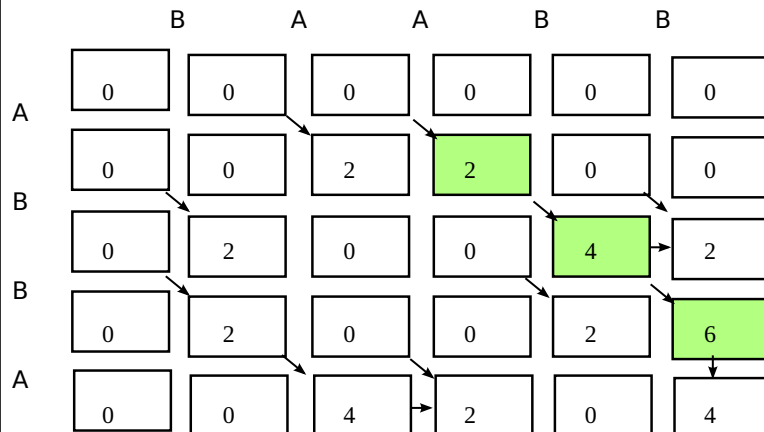
Poniżej przedstawiono macierz dla algorytmu, który dostarcza uliniowanie lokalne (algorytm Smitha-Watermana) dwóch sekwencji. Macierz podobieństwa pokazano obok. Stosujemy liniową karę za przerwę.

	A	B	C
A	2	-2	-2
B	-2	2	-2
C	-2	-2	2



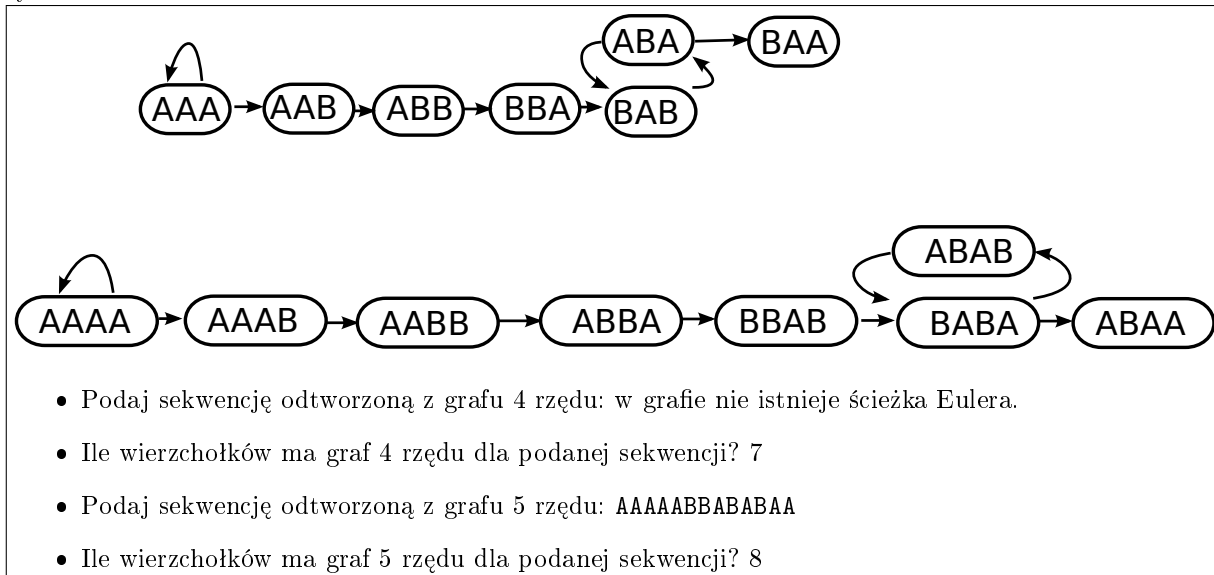
Odpowiedzi:

- Algorytm uruchomiono dla sekwencji ABBA oraz BAABB:
- Kara za przerwę (parametr d) wynosi -2
- Maksymalnie podobne pod-sekwencje to ABB (3 pierwsze symbole z ABBA) i ABB (3 ostatnie symbole z BAABB)



Zadanie 2 (8 pkt)

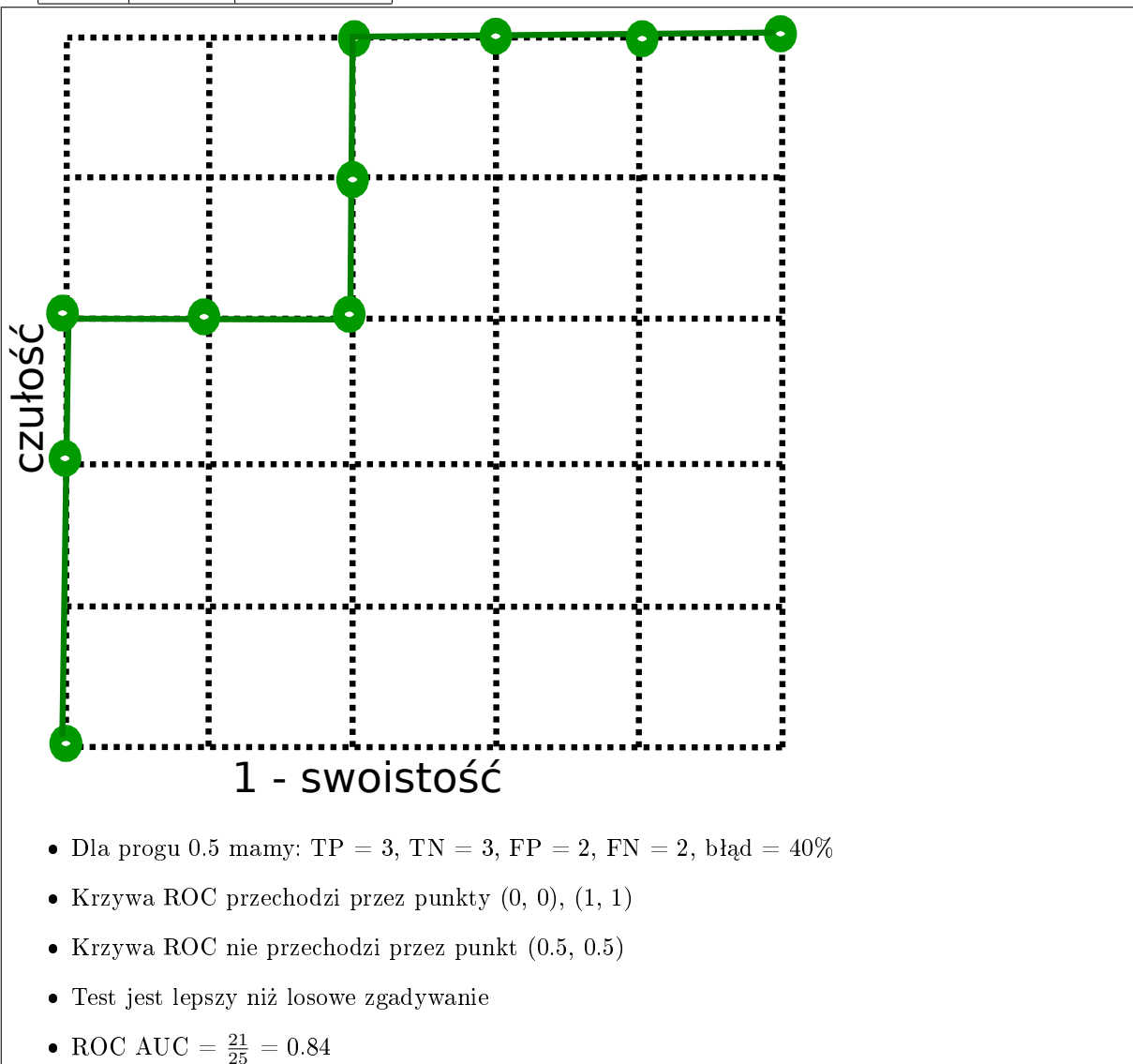
Sekwencję AAAAAABBABABAA użyto do budowy grafu de Bruijina (bez wag w krawędziach) 4 rzędu (w wierzchołkach są sekwencje o długości 3 symboli) oraz 5 rzędu (w wierzchołkach są sekwencje o długości 4 symboli).



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=+
ATGTTGTAAGAGGAGACCACCGCGCTC
```

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

Odpowiedzi:

- dokonaj translacji sekwencji referencyjnej DNA na sekwencję aminokwasów
 - Sekwencja referencyjna AA: MCEEETAL
 - Sekwencja pacjenta AA: MGE-ETTAL
- Znajdź warianty SNV na poziomie sekwencji DNA i zapisz je w formacie: CHR:POS_REF>ALT, np: chr1:1000000_A>T

REF

ATG TGT GAA GAG GAG ACC ACC GCG CTC

ALT

ATG GGT GAA TAG GAG ACC ACC GCA CTC

start = 73901312

Zmiany na poziomie DNA:

chr2:73901315_T>G

chr2:73901321_G>T

chr2:73901335_G>A

- Określ efekt każdego wariantu jak będzie on miał dla sekwencji białkowej. W szczególności podaj liczbę zmian synonimicznych i niesynonimicznych.

Zmiany na poziomie AA: MGEETAL
MCE-ETTAL

- chr2:73901315_T>G - zmiana typu missense (zmiana drugiego aminokwasu (G → C))
- chr2:73901321_G>T - zmiana typu stopgain (E → -)
- chr2:73901335_G>A - zmiana typu synonymous (A → A)

Występują 2 warianty niesynonimiczne i jeden synonimiczny.

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

