

Imię i nazwisko:

Zadanie 1 (8 pkt)

Poniżej przedstawiono macierz dla algorytmu, który dostarcza uliniwienia globalnego (algorytm Needlemana-Wunscha) dwóch sekwencji. Macierz podobieństwa pokazano obok. Stosujemy liniową karę za przerwę, $\gamma(n) = n * d, d = -4$.

	A	B
A	2	-3
B	-3	2

	A	A	B	A	A	
A	0	-4	-8	-12	-16	-20
B	-4	2	-2	-6	-10	-14
A	-8	-2	-1	0	-4	-8
B	-12	-6	0	-4	2	-2
B	-16	-10	-4	2	-2	-1

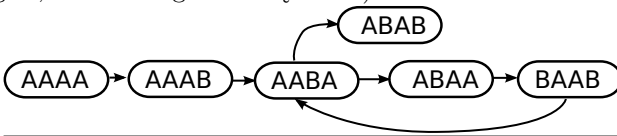
Pytania:

- algorytm uruchomiono dla sekwencji AABAA i ABAB
- rozwiązań optymalnych jest 2 (na rysunku brakuje jednej strzałki)
- rozwiązania optymalne to:

AABAA AABAA
 -ABAB A-BAB

Zadanie 2 (8 pkt)

Dla zestawu odczytów zbudowano graf de Bruijna 5 rzędu pokazany na rysunku (nie jest to multi-graf, ani A-graf, nie ma wag w krawędziach).



Pytania:

- czy istnieje rozwiązanie (ścieżka Eulera)? **tak**
- jeżeli tak, podaj sekwencję odpowiadającą jednemu z rozwiązań: **AAAABAABAB**
- zbuduj graf de Bruijna 4 rzędu (bez wag w krawędziach). Ile ten graf ma wierzchołków: **5**
- czy istnieje rozwiązanie (ścieżka Eulera)? **nie**
- jeżeli tak, to podaj sekwencję odpowiadającą jednemu z rozwiązań: -

Zadanie 5 (8 pkt)

Cząsteczka jest biopolimerem i może być reprezentowana przez napis nad alfabetem { A, B }. Obserwujemy oddziaływanie poszczególnych symboli z naszą sondą i oznaczamy je jako '+' lub '-'. Chcemy opisać naszą cząsteczkę oraz oddziaływanie za pomocą ukrytego modelu Markowa. Stanem będzie symbol, więc $Q = \{A, B\}$, obserwacją oddziaływanie, więc $V = \{+, -\}$. Dla cząsteczki AAAAAAAAAA obserwujemy oddziaływanie $+ - + + + - - - -$, dla cząsteczki BBBB obserwujemy $+ - + +$. Nasze sekwencje zaczynają się od z takim samym prawdopodobieństwem od A i od B, więc $P_A = \frac{1}{2}, P_B = \frac{1}{2}$. Znamy macierz przejść:

	A	B
A	$\frac{4}{5}$	$\frac{1}{5}$
B	$\frac{1}{5}$	$\frac{4}{5}$

Pytania:

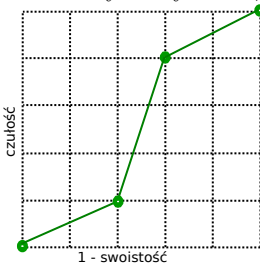
- podaj macierz emisji:

	+	-
A	$\frac{1}{2}$	$\frac{1}{2}$
B	$\frac{3}{4}$	$\frac{1}{4}$

- podaj najbardziej prawdopodobną sekwencję cząsteczki dla oddziaływania + + + **BBB**

Zadanie 4 (8 pkt)

Opracowano test binarny X, dla którego krzywa ROC jest pokazana na rysunku. Test zwraca wynik 'tak' lub 'nie'. Używamy testu, gdzie obiekty mają w 50% stan 'tak' i w 50% stan 'nie'.



Pytania:

- czy test może zwracać zawsze wynik 'nie' ? **tak**
- jaka jest wtedy czułość? **0**
- jaka jest czułość dla swoistości 0.4 ? **0.8**
- Jaki będzie błąd przy czułości 0.2? **0.6**
- Jaki będzie błąd przy czułości 0.8? **0.4**

Zadanie 5 (8 pkt)

Mamy dane pokazane poniżej i zakładamy pierwszą ramkę odczytu (pierwszy kodon odpowiada pierwszemu aminokwasowi).

- a) fragment sekwencji kodującej genu CLP1 w genomie referencyjnym:

```
>hg38_dna range=chr11:57659477-57659493 5'pad=0 3'pad=0 strand=+  
ATGGGAGAAGAGGCT
```

- b) sekwencję pacjenta:

```
> patient1  
ATGGGAGTAGAGGCC
```

Tablica kodowania aminokwasów:

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

Polecenia:

- Dokonaj translacji sekwencji referencyjnej DNA na sekwencję aminokwasów

MGEEA

- Dokonaj translacji sekwencji DNA pacjenta na sekwencję aminokwasów

MGVEA

- Znajdź warianty SNV na poziomie sekwencji DNA i zapisz je w formacie:

- CHR:POS_REF>ALT, np: chr11:57659477_A>T

chr11:57659484 A>T

chr11:57659491 T>C

- Określ efekt każdego wariantu jak będzie on miał dla sekwencji białkowej, nadając im jedną z 4 etykiet: 'missense', 'synonymous', 'stopgain', 'stoploss'

chr11:57659484 - zmiana typu missense

chr11:57659491 - zmiana typu synonymous