

Metody Bioinformatyki, część informatyczna, 4 lutego 2019, przykładowe rozwiązania

Imię i nazwisko:

Zadanie 1 (9 pkt)

Uzupełnij macierz dla algorytmu, który bada podobieństwo globalne (algorytm Needlemana-Wunscha). Podaj najlepsze rozwiązania dla sekwencji **AABBA** oraz profilu:

A	0.8	1	0	0.4
B	0.2	0	1	0.6

	A	B
A	5	-10
B	-10	5

Macierz podobieństwa pokazano obok. Stosujemy liniową karę za przerwę, $d = -6$.

	A	A	B	B	A	
0	0	-6	-12	-18	-24	-30
1	-6	2	-4	-10	-16	-22
2	-12	-1	7	1	-5	-11
3	-18	-7	1	12	6	0
4	-24	-13	-5	6	11	5

Rozwiązania:

A	A	B	B	A
0	1	2	3	-

Ilość rozwiązań

Zadanie 2 (7 pkt)

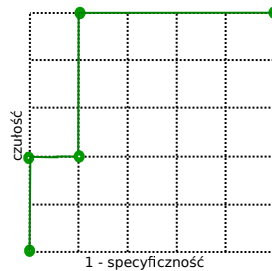
Opracowano test, który zwraca prawdopodobieństwo wystąpienia choroby, jego wyniki są przedstawione w tabeli. Podaj macierz pomyłek, zakładając, że traktujemy wynik testu powyżej 0.45 jako pozytywny (osoba chora), następnie podaj tę macierz dla progów 0.65 oraz 0.75. Narysuj krzywą ROC, a na niej 5 punktów (dla progów: 0.0, 0.45, 0.65, 0.75, 1.0).

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

Próg 0.45:	TP = 5	FP = 1
	FN = 0	TN = 4

Próg 0.65:	TP = 2	FP = 1
	FN = 3	TN = 4

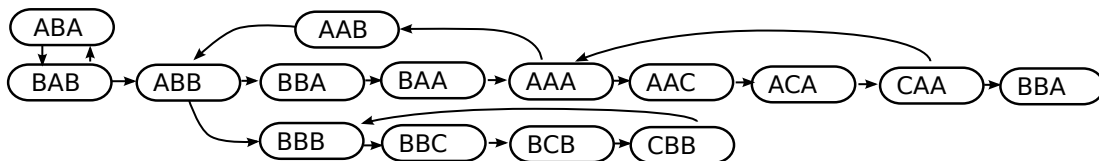
Próg 0.75:	TP = 2	FP = 0
	FN = 3	TN = 5



Zadanie 3 (8 pkt)

Dla zestawu odczytów: **ABBBCBBB, BAAACAAAB, BABBAA, BABAB, AABB**, zbuduj multi-graf de Bruijna 4 rzędu (w wierzchołkach są sekwencje o długości 3 symboli, krawędzie przechowują informację, ile razy krawędź ma być dodana do ścieżki wynikowej).

Podaje sekwencję wynikową. Jeżeli istnieje więcej niż jedno rozwiązanie, podaj dowolne z nich.

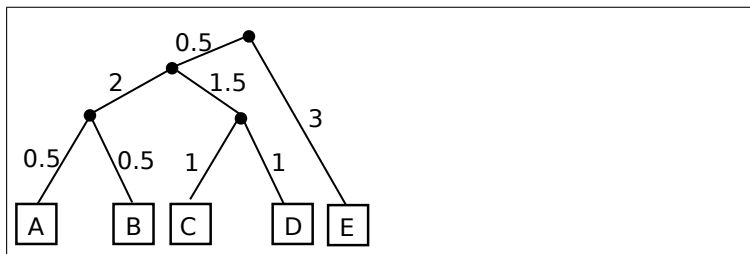


Sekwencja: **BABABBAACAAABBBBCBBB**

Zadanie 4 (7 pkt)

Podaj drzewo filogenetyczne dla taksonów A, B, C, D, E dla pokazanej macierzy odległości. Wykorzystaj metodę grupowania parami ze średnią arytmetyczną (UPGMA).

	B	C	D	E
A	1	6	4	5
B		4	6	7
C			2	7
D				5



Zadanie 5 (9 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 *OOO*.

$$Q = \{A, B, C\}$$

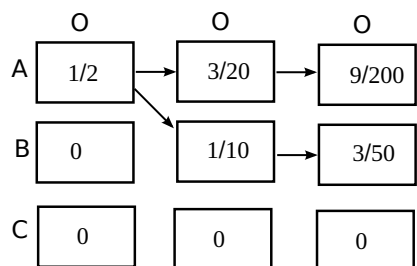
$$V = \{O, R\}$$

$$P_A = 1, P_B = 0$$

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

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

ABB



Notatki lub uwagi do prowadzącego