

Metody Bioinformatyki, część informatyczna, 28 czerwca 2016

Imię i nazwisko:

y

czas: 45 min

Zadanie 1 (8 pkt)

Uzupełnij macierz dla algorytmu, który bada podobieństwo lokalne (algorytm Smitha-Watermana). Podaj najlepsze rozwiązania dla sekwencji **BBABB** oraz **ABBA**. Macierz podobieństwa pokazano obok. Stosujemy liniową karę za przerwę, $\gamma(n) = n * d, d = -3$.

	A	B
A	2	-1
B	-1	2

	B	B	A	B	B
A					
B					
B					
A					
A					

Rozwiązania:

Ilość rozwiązań

Zadanie 2 (8 pkt)

Opracowano test X, który zwraca prawdopodobieństwo wystąpienia choroby. Wyniki testu są przedstawione w tabeli. Podaj macierz pomyłek, zakładając, że traktujemy wynik testu powyżej 0.35 jako pozytywny (osoba chora), następnie podaj tę macierz dla progów 0.55 oraz 0.75. Narysuj krzywą ROC, a na niej 3 punkty (dla progów: 0.35, 0.55, 0.75).

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

Próg 0.35:

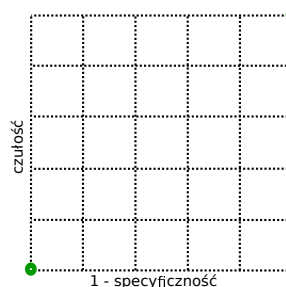
TP =	FP =
FN =	TN =

Próg 0.55:

TP =	FP =
FN =	TN =

Próg 0.75:

TP =	FP =
FN =	TN =



Zadanie 3 (5 pkt)

Podaj drzewo filogenetyczne dla 4 taksonów (A, B, C, D), macierz odległości pokazano poniżej. Wykorzystaj metodę grupowania parami ze średnią arytmetyczną (UPGMA).

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

Zadanie 4 (3 pkt)

Obserwujemy dwa loci A i B , częstości haplotypów są następujące: $P(A_1B_1) = P(A_2B_2) = 0.3$, $P(A_1B_2) = P(A_2B_1) = P(A_2B_2) = P(A_3B_1) = P(A_3B_2) = 0.1$. Podaj genotyp, który występuje najczęściej oraz prawdopodobieństwo jego wystąpienia.

Zadanie 5 (9 pkt)

Posługujemy się trzema monetami, obserwując sekwencje rzutów (orły i reszki). Zakładając, że przedstawione doświadczenie jest opisywane ukrytym modelem Markowa przedstawionym obok, podaj prawdopodobieństwo uzyskania sekwencji RR .

- $Q = \{U, N1, N2\}$
- $V = \{O, R\}$
- $P_U = 0, P_{N1} = 1, P_{N2} = 0$

	U	N1	N2
U	$\frac{2}{3}$	$\frac{1}{3}$	0
N1	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
N2	0	$\frac{1}{3}$	$\frac{2}{3}$

	O	R
U	$\frac{1}{2}$	$\frac{1}{2}$
N1	$\frac{1}{3}$	$\frac{2}{3}$
N2	0	1

Zadanie 6 (7pkt)

Dla zestawu odczytów: AAABA, AAAABB, ABBBBBBBB, BBBBAAA, ABAAAAA, zbuduj graf de Bruijna 5 rzędu (w wierzchołkach są sekwencje o długości 4 symboli).

Czy istnieje ścieżka Eulera? (tak/nie)?

Podaje sekwencję wynikową, która jest najdłuższą ścieżką w grafie (w szczególności jest to ścieżka Eulera, jeżeli graf taką posiada). Jeżeli istnieje więcej niż jedno rozwiązanie, podaj dowolne z nich.

Sekwencja:

Notatki lub uwagi do prowadzącego