

Metody Bioinformatyki, część informatyczna, 8 lutego 2016

Imię i nazwisko: u czas: 50 min

Zadanie 1 (8 pkt)

Uzupełnij macierz dla algorytmu, który bada podobieństwo globalne (algorytm Needlemana-Wunscha). Podaj najlepsze rozwiązania dla sekwencji **BABAB** oraz **ABBB**. Macierz podobieństwa pokazano obok. Stosujemy liniową karę za przerwę, $\gamma(n) = n * d, d = -1$.

	A	B
A	2	-2
B	-2	2

	B	A	B	A	B
A					
B					
B					
B					
B					

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 0.45 jako pozytywny (osoba chora), następnie podaj tę macierz dla progu 0.55 oraz progu 0.75. Narysuj 3 punkty na krzywej ROC (0.45, 0.55, 0.75). Narysuj krzywą ROC.

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

Próg 0.45:

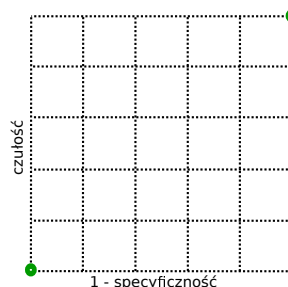
TP =	FP =
FN =	TN =

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).

	A	B	C	D
A	0	3	2	3
B		0	5	5
C			0	7
D				0

Zadanie 4 (3 pkt)

Dla populacji obserwujemy dwa loci A i B , każde ma dwa warianty: A i a , oraz B i b . Podaj genotyp, który występuje najczęściej, jeżeli częstości występowania haplotypów są następujące: $P(ab) = 0.3$, $P(Ab) = 0.2$, $P(aB) = 0.2$, $P(AB) = 0.3$. Zakładamy równowagę Hardy'ego-Weinberga dla obserwowanej populacji.

Zadanie 5 (9 pkt)

Posługujemy się trzema monetami, jedna jest uczciwa, obserwując sekwencje rzutów (orły i reszki). Zakładając, że przedstawione doświadczenie jest opisywane ukrytym modelem Markowa przedstawionym obok, podaj najbardziej prawdopodobną sekwencję stanów (sekwencję użytych monet), jeżeli wynikiem doświadczenia jest sekwencja $ORRR$.

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

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

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

Zadanie 6 (7pkt)

Dla zestawu odczytów: AAAABA, ABBBBAAA, AAABA, AABAAA, BAAAB, AABBB 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