

Imię i nazwisko:

Zadanie 1 (8 pkt)

Uzupełnij macierz dla algorytmu, który dostarcza uliniowienia globalnego (algorytm Needlemana-Wunscha). Podaj najlepsze rozwiązania dla sekwencji **BCACB** oraz **CABA**. Macierz podobieństwa pokazano obok. Stosujemy liniową karę za przerwę, $\gamma(n) = n * d, d = -2$.

	A	B	C
A	2	-5	-5
B	-5	2	0
C	-5	0	2

	B	C	A	C	B
C					
A					
B					
A					
A					

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 progu 0.45 oraz 0.55. Narysuj krzywą ROC i krzywą PR.

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

Próg 0.35:	TP =	FP =
	FN =	TN =
Próg 0.45:	TP =	FP =
	FN =	TN =
Próg 0.55:	TP =	FP =
	FN =	TN =

czułość

1 - specyficzność

precyzja

czułość

Zadanie 3 (8 pkt)

Dla zestawu odczytów: AAAAAA, BBAAA, BABABAB i AAABAB 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. Podaj sekwencję wynikową. Jeżeli istnieje więcej niż jedno rozwiązanie, podaj dowolne z nich.

Sekwencja:

Zadanie 4 (8 pkt)

Podaj drzewo filogenetyczne dla taksonów A, B, C, D dla pokazanej macierzy odległości. Wykorzystaj metodę grupowania parami ze średnią arytmetyczną (UPGMA). Dla tej samej macierzy podaj dwie grupy zwracane przez algorytm grupowania hierarchicznego, gdy odległość między grupami to pełne wiązanie (maksymalna odległość między elementami w grupie).

	B	C	D
A	3	2	5
B		5	6
C			7

Grupa 1:

Grupa 2:

Zadanie 5 (8 pkt)

Cząsteczka jest biopolimerem i może być reprezentowana przez napis nad alfabetem { A, B, C }. Obserwujemy oddziaływanie poszczególnych symboli z naszą sondą i oznaczamy je jako '+' lub '-'. Podaj najbardziej prawdopodobną sekwencję cząsteczki, dla obserwacji + - -, jeżeli zależność między kolejnymi symbolami oraz prawdopodobieństwo wystąpienia oddziaływania opiszemy ukrytym modelem Markowa przedstawionym niżej.

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

$$V = \{+, -\}$$

$$P_A = 1, P_B = 0, P_C = 0$$

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

	+	-
A	$\frac{1}{2}$	$\frac{1}{2}$
B	1	0
C	0	1

Sekwencja cząsteczki:

	+	-	-
A	<input style="width: 50px; height: 30px;" type="text"/>	<input style="width: 50px; height: 30px;" type="text"/>	<input style="width: 50px; height: 30px;" type="text"/>
B	<input style="width: 50px; height: 30px;" type="text"/>	<input style="width: 50px; height: 30px;" type="text"/>	<input style="width: 50px; height: 30px;" type="text"/>
C	<input style="width: 50px; height: 30px;" type="text"/>	<input style="width: 50px; height: 30px;" type="text"/>	<input style="width: 50px; height: 30px;" type="text"/>

Notatki lub uwagi do prowadzącego