

Imię i nazwisko:

Zadanie 1 (8 pkt)

Uzupełnij macierz dla algorytmu, który bada podobieństwo lokalne globalne (algorytm Needlemana-Wunscha). Podaj najlepsze rozwiązania dla sekwencji **BBABB** oraz **BABA**. Macierz podobieństwa pokazano obok. Stosujemy liniową karę za przerwę, $\gamma(n) = n * d, d = -2$.

	A	B
A	2	-5
B	-5	2

	B	B	A	B	B	
	0	-2	-4	-6	-8	-10
B	-2	2	0	-2	-4	-6
A	-4	0	-2	2	0	-2
B	-6	-2	2	0	4	2
A	-8	-4	0	4	2	0

Rozwiązania:

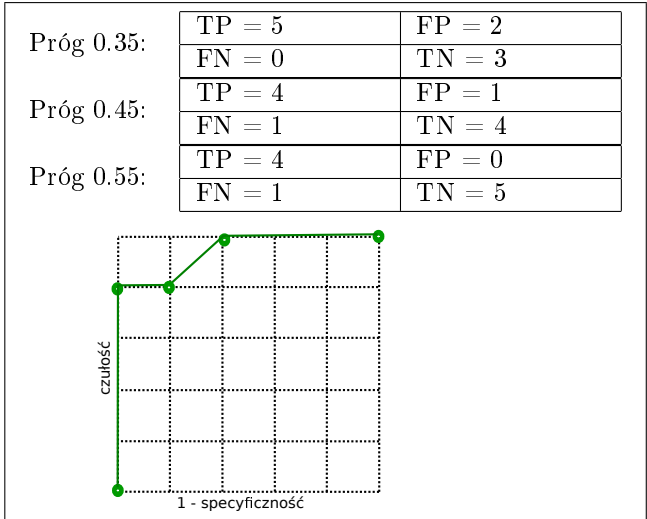
B-BABB	BBAB-B	BBAB-B	
BABA--	-BABA-	B-ABA-	
BBABB-	BBABB-	BBABB-	BBABB-
-BAB-A	-BA-BA	B-AB-A	B-A-BA

Ilość rozwiązań 7

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 0.45 oraz 0.55. Narysuj krzywą ROC, a na niej 5 punktów (dla progów: 0.0, 0.35, 0.45, 0.55, 1.0).

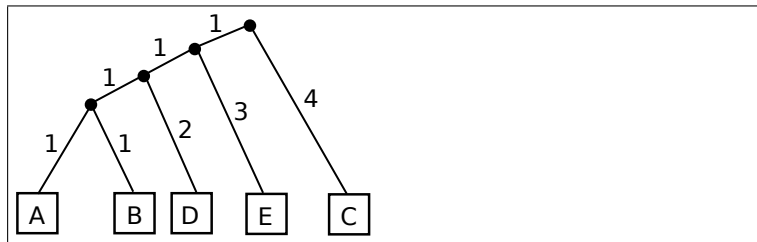
osoba	stan	wynik testu
A	zdrowa	0.1
B	zdrowa	0.2
C	zdrowa	0.3
D	chora	0.4
E	zdrowa	0.4
F	zdrowa	0.5
G	chora	0.6
H	chora	0.7
I	chora	0.8
J	chora	0.9



Zadanie 3 (7 pkt)

Podaj drzewo filogenetyczne dla 5 taksonów (A, B, C, D, E), macierz odległości pokazano poniżej. Wykorzystaj metodę grupowania parami ze średnią arytmetyczną (UPGMA).

	B	C	D	E
A	2	7	3	5
B		9	5	7
C			8	8
D				6



Zadanie 4 (9 pkt)

Posługujemy się czterema monetami, obserwując sekwencje rzutów (orły i reszki). Zakładając, że przedstawione doświadczenie jest opisywane ukrytym modelem Markowa przedstawionym obok, podaj wszystkie najbardziej prawdopodobne sekwencje stanów (sekwencję użytych monet), gdy obserwujemy *OOOR*.

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

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

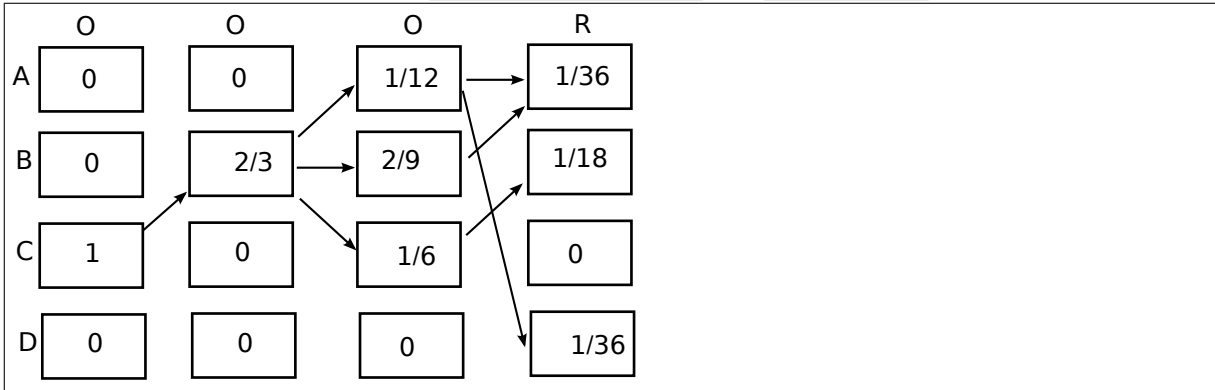
$$P_A = 0, P_B = 0$$

$$P_C = 1, P_D = 0$$

	A	B	C	D
A	$\frac{2}{3}$	0	0	$\frac{1}{3}$
B	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	0
C	0	1	0	0
D	0	1	0	0

	O	R
A	$\frac{1}{2}$	$\frac{1}{2}$
B	$\frac{2}{3}$	$\frac{1}{3}$
C	1	0
D	0	1

C B C B

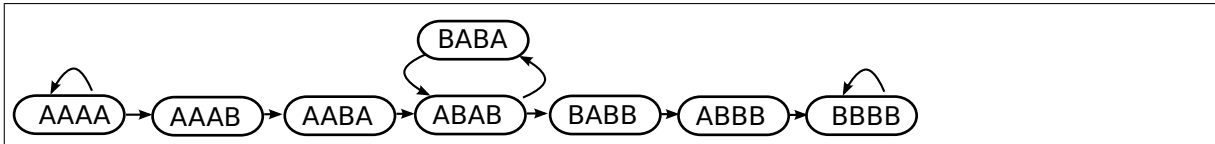


Zadanie 5 (8 pkt)

Dla zestawu odczytów: ABBBBB, AAAAAAB, AAABABA, BABABBB, zbuduj graf de Bruijna 5 rzędu (w wierzchołkach są sekwencje o długości 4 symboli).

Czy istnieje ścieżka Eulera? **tak**

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: AAAAABABABBBBB