

**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,  $\gamma(n) = n * d, d = -3$ .

	<b>A</b>	<b>B</b>
<b>A</b>	2	-1
<b>B</b>	-1	2

		<b>B</b>	<b>B</b>	<b>A</b>	<b>B</b>	<b>B</b>
		0	0	0	0	0
<b>A</b>		0	0	0	2	0
<b>B</b>		0	2	2	0	4
<b>B</b>		0	2	4	1	2
<b>A</b>		0	0	1	6	3

Rozwiązania:

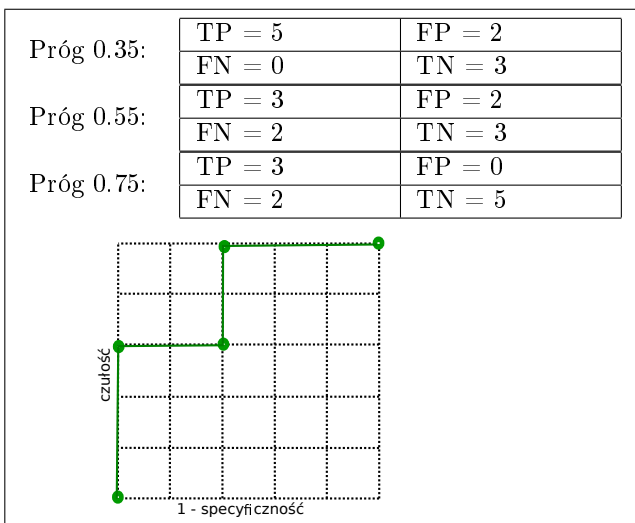
BBA	ABB
BBA	ABB

Ilość rozwiązań **2**

**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 0.55 oraz 0.75. Narysuj krzywą ROC, a na niej 3 punkty (dla progów: 0.35, 0.55, 0.75).

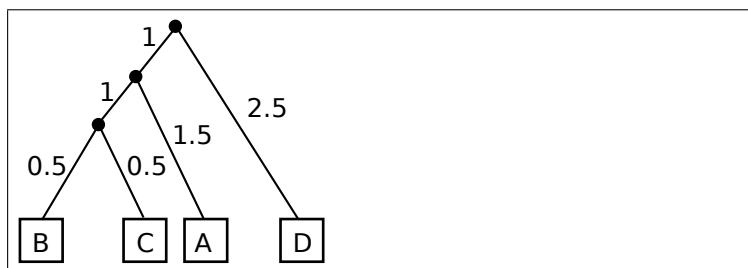
osoba	stan	wynik testu
A	zdrowa	0.1
B	zdrowa	0.2
C	zdrowa	0.3
D	chora	0.4
E	chora	0.5
F	zdrowa	0.6
G	zdrowa	0.7
H	chora	0.8
I	chora	0.9
J	chora	0.9



**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>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	2	4	5
<b>B</b>		1	4
<b>C</b>			6



**Zadanie 4 (3 pkt)**

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

$$P(A_1A_2B_1B_2) = 2 * P(A_1B_1) * P(A_2B_2) + 2 * P(A_1B_2) * P(A_2B_1) = 0.2$$

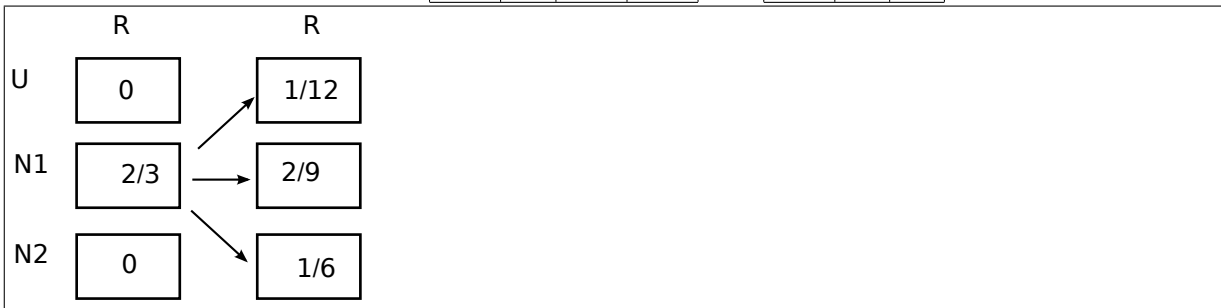
**Zadanie 5 (9 pkt)**

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  $RR$ .

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

	<b>U</b>	<b>N1</b>	<b>N2</b>		<b>O</b>	<b>R</b>
<b>U</b>	$\frac{2}{3}$	$\frac{1}{3}$	0	<b>U</b>	$\frac{1}{2}$	$\frac{1}{2}$
<b>N1</b>	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	<b>N1</b>	$\frac{1}{3}$	$\frac{2}{3}$
<b>N2</b>	0	$\frac{1}{3}$	$\frac{2}{3}$	<b>N2</b>	0	1

$$\frac{17}{36} \approx 0.472$$

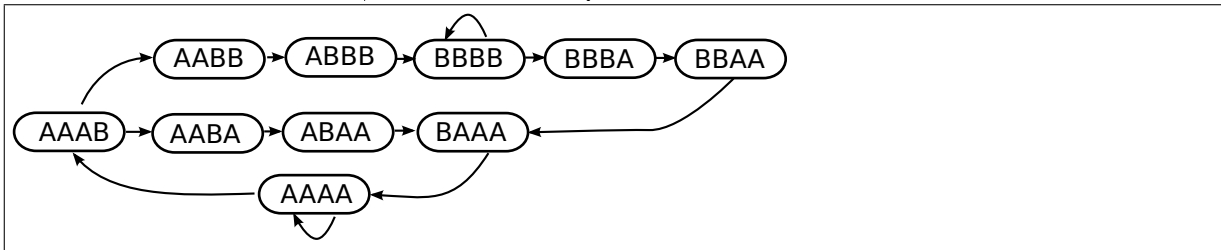


**Zadanie 6 (7pkt)**

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

Czy istnieje ścieżka Eulera? (tak/nie)? **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.



Seqwencja (jedna z kilku): **AAABAAAAABBBBBBAAA**

Notatki lub uwagi do prowadzącego