

Metody Bioinformatyki, część informatyczna, 14 czerwca 2016

Imię i nazwisko:

V

czas: 45 min

Zadanie 1 (8 pkt)

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

| | | |
|---|----|----|
| | A | B |
| A | 3 | -1 |
| B | -1 | 3 |

| | | | | | | |
|---|----|----|----|----|----|-----|
| | B | B | B | A | A | |
| A | 0 | -2 | -4 | -6 | -8 | -10 |
| B | -2 | -1 | -3 | -5 | -3 | -5 |
| B | -4 | 1 | 2 | 0 | -2 | -4 |
| B | -6 | -1 | 4 | 5 | 3 | 1 |
| B | -8 | -3 | 2 | 7 | 5 | 3 |

Rozwiązania:

-BBBAA
ABBB--

Ilość rozwiązań: 1

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 0.65 oraz 0.75. Narysuj krzywą ROC, a na niej 3 punkty (dla progów: 0.35, 0.65, 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 | chora | 0.6 |
| G | zdrowa | 0.7 |
| H | chora | 0.8 |
| I | chora | 0.9 |
| J | zdrowa | 0.9 |

Próg 0.35:

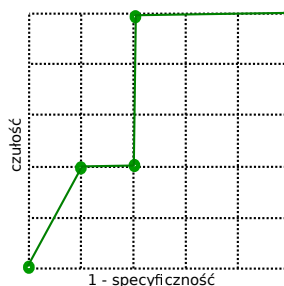
| | |
|--------|--------|
| TP = 5 | FP = 2 |
| FN = 0 | TN = 3 |

Próg 0.65:

| | |
|--------|--------|
| TP = 2 | FP = 2 |
| FN = 3 | TN = 3 |

Próg 0.75:

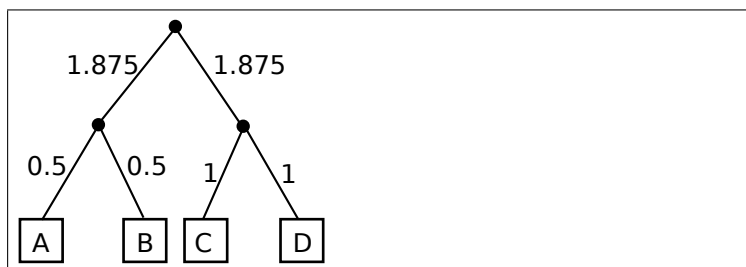
| | |
|--------|--------|
| TP = 2 | FP = 1 |
| FN = 3 | TN = 4 |



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 | 1 | 4 | 3 |
| B | | 5 | 3 |
| C | | | 2 |



Zadanie 4 (3 pkt)

Obserwujemy dwa loci A i B , każde ma trzy warianty: A_1, A_2, A_3 i B_1, B_2 i B_3 . Podaj częstość występowania genotypu A_1B_1 , jeżeli częstości występowania wariantów są następujące: $P(A_1) = P(A_2) = 0.3$, $P(A_3) = 0.4$, $P(B_1) = 0.2$, $P(B_2) = 0.3$, $P(B_3) = 0.5$. Loci nie są sprzężone.

$$P(A_1B_1) = P(A_1)^2P(B_1)^2 = 0.036$$

Zadanie 5 (9 pkt)

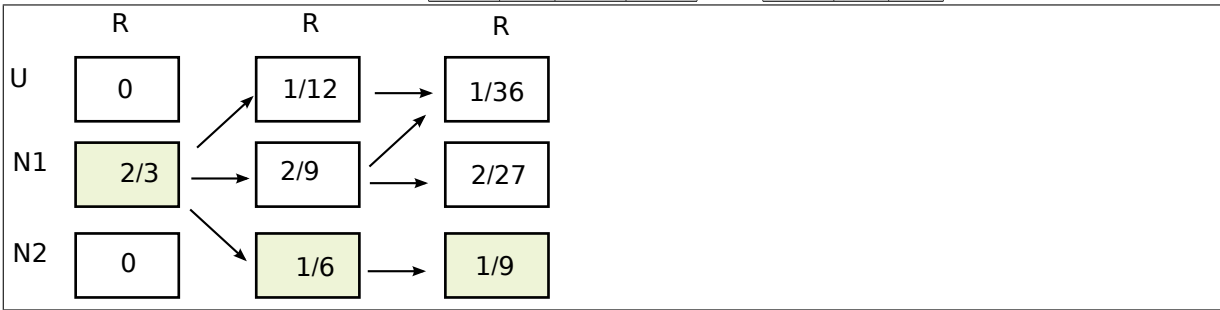
Posługujemy się trzema monetami, obserwując sekwencje rzutów (orły i reszki). Zakładając, że przedstawione doświadczenie jest opisywane ukrytym modelem Markowa przedstawionym obok, podaj najbardziej prawdopodobną sekwencję stanów (sekwencję użytych monet), jeżeli wynikiem doświadczenia jest sekwencja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

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

| | | | |
|----|---------------|---------------|---------------|
| | 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 |

N1 N2 N2

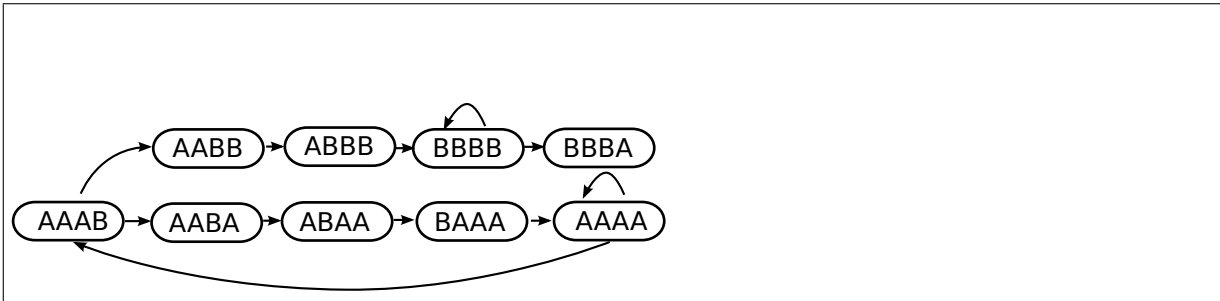


Zadanie 6 (7pkt)

Dla zestawu odczytów: AAABA, AAAABB, ABBBBB, BBBBA, AABAAAA, 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.



Seqwencja: AAABAAAAABBBBBBA

Notatki lub uwagi do prowadzącego