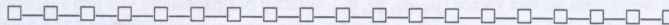


TENTAMEN INTRODUCTION TO BIOMEDICAL COMPUTING

4-2-2011



Voorzie de in te leveren bladen van je naam, en nummer ze. Schrijf op het eerste blad het aantal ingeleverde bladen. Bij elk van de 4 opgaven is het maximale aantal voor deze opgave te behalen punten vermeld. Je krijgt 10 punten gratis. Succes!

Opgave 1 Sequence alignment (25 pt)

Hieronder is de pseudo-code van het Needleman-Wunsch algoritme weergegeven.

```

1: INPUT: twee sequenties  $s = a_1 a_2 \dots a_n$  en  $t = b_1 b_2 \dots b_m$ ; kostenfunctie  $w$  en gap penalty  $g$ 
2: OUTPUT: edit distance matrix  $D$ 
3: for  $i = 0$  to  $n$  do
4:    $D(i, 0) \leftarrow g \cdot i$ 
5: end for
6: for  $j = 0$  to  $m$  do
7:    $D(0, j) \leftarrow g \cdot j$ 
8: end for
9: for  $i = 1$  to  $n$  do
10:  for  $j = 1$  to  $m$  do
11:   Match  $\leftarrow D(i-1, j-1) + w(a_i, b_j)$ 
12:   Delete  $\leftarrow D(i-1, j) + g$ 
13:   Insert  $\leftarrow D(i, j-1) + g$ 
14:    $D(i, j) \leftarrow \min(\text{Match}, \text{Insert}, \text{Delete})$ 
15:  end for
16: end for
    
```

- Documenteer de body in de regels 3-16 van dit algoritme, d.w.z. beschrijf voor elke regel en/of bijeen horende reeks van regels welke operatie wordt uitgevoerd en met welk doel.
- Teken de dotplot van de DNA-sequenties: $s = \text{T C A G}$ en $t = \text{A T C A}$.
- In veel DNA-sequenties komen repeterende deelsequenties voor. Hoe uiten zulke repetities zich in een dotplot van de DNA-sequentie met zichzelf?
- Bekijk nu het Needleman-Wunsch algoritme voor de recursieve berekening van de edit distance matrix D van de sequenties s en t zoals aangegeven onder b. De kostenfunctie, inclusief gap penalties, heeft de volgende vorm:

	A	T	G	C	-
A	0	2	1	2	3
T	2	0	2	1	3
G	1	2	0	2	3
C	2	1	2	0	3
-	3	3	3	3	

Deze matrix is gebaseerd op het biologische verschijnsel dat mutaties $A \leftrightarrow G$ en $T \leftrightarrow C$ vaker voorkomen dan $(A \text{ of } G) \leftrightarrow (T \text{ of } C)$.

Bereken de edit distance matrix D , inclusief pointers naar de predecessor(s).

- Geef het aantal *trace-back* paden in de matrix D , beginnend rechtsonder en eindigend linksboven.
- Geef de bijbehorende optimale alignments en hun edit distance.

(Ga verder op de volgende pagina)

Opgave 2 Booleaanse netwerken (25 pt)

Neem een booleaans netwerk met $N = 4$ en K onbekend. De toestand van het hele netwerk op elk moment kunnen we weergeven als een vector (g_4, g_3, g_2, g_1) , waarbij g_i de toestand van gen i is, voor $i = 1, 2, 3, 4$, waarbij $g_i = 0$ of $g_i = 1$. Er zijn dus $2^4 = 16$ toestanden, genummerd 0000 t/m 1111 binair, ofwel 0 t/m 15 decimaal. Ga uit van de volgende toestandsovergangen:

t_n	t_{n+1}
0	→ 8
1	→ 9
2	→ 9
3	→ 15
4	→ 14
5	→ 13
6	→ 13
7	→ 9
8	→ 9
9	→ 10
10	→ 8
11	→ 9
12	→ 13
13	→ 14
14	→ 13
15	→ 12

- Maak een diagram van deze overgangen. Hoeveel limit cycles zijn er, hoe zien de toestandsovergangen van elke limit cycle er uit, wat is de periode en wat zijn de "basins of attraction"? Geef ook de langste "run-in" lengte (duur voordat de limit cycle bereikt wordt).
- Gegeven dat het systeem in een limit cycle toestand zit, wat is het kleinste aantal bits dat in de toestand moet worden gewijzigd om naar een andere limit cycle te gaan (m.a.w. naar de basin of attraction van een andere limit cycle)?
- Is er een node (gen) met een waarheidstabel die leidt tot bevroren gedrag? Zo ja welke node(s), en waarom; zo nee, waarom niet?

Opgave 3 Tomografische reconstructie (25 pt)

Hieronder is de pseudo-code van het Kaczmarz algoritme (zonder constraints) weergegeven.

```

1: INPUT: projectiewaarden (ray sums)  $p_1, p_2, \dots, p_M$  met gewichtsvectoren  $\vec{w}_1, \vec{w}_2, \dots, \vec{w}_M$ 
2: INPUT: parameters  $\text{MAX\_ITER} \geq 1, \epsilon > 0$ 
3: OUTPUT: vector representatie  $\vec{f} = (f_1, f_2, \dots, f_N)$  van het gereconstrueerde plaatje.
4: Initialiseer  $\vec{f}^{(0)}$ 
5: for  $k = 1$  to  $\text{MAX\_ITER}$  do
6:   for  $i = 1$  to  $M$  do
7:     bereken  $\vec{f}^{(i)}$  uit  $\vec{f}^{(i-1)}$ 
8:   end for
9:   if  $\delta(\vec{f}^{(M)}, \vec{f}^{(0)}) < \epsilon$  then
10:    break
11:   end if
12:    $\vec{f}^{(0)} \leftarrow \vec{f}^{(M)}$ 
13: end for

```

(Ga verder op de volgende pagina)

We werken met een 2×2 -plaatje waarvan we de pixelwaarden kennen, zoals weergegeven in de volgende figuur:

4	4
4	8

We nemen aan dat MAX_ITER heel groot en ϵ heel klein is.

- Bereken de vier projectiewaarden over de twee rijen en de twee kolommen. We noemen dit *projectiemethode 1*.
- Pas nu het Kaczmarz algoritme toe en teken de toestand van het plaatje (d.w.z. de vier pixelwaarden) op dezelfde manier als in bovenstaande figuur, voor $k = 1$ na uitvoering van regel 7 voor $i = 1$, $i = 2$, $i = 3$ en $i = 4$. Initialiseer alle pixelwaarden op 0.
- De oplossing zal in de volgende iteratie niet meer veranderen, en het algoritme termineert voor $k = 2$. Leg uit waarom.
- Is de gevonden oplossing correct (d.w.z., zijn de pixelwaarden zoals in het plaatje waarvan we zijn uitgegaan)? Zo nee, geef hiervoor een verklaring.
Hint: bekijk de nulruimte bij projectiemethode 1.
- Nu passen we een tweede manier toe om vier projecties te berekenen, namelijk:
 - Projectie 1: over rij 1
 - Projectie 2: over kolom 1
 - Projectie 3: over de noordwestelijke diagonaal (pixel linksboven \rightarrow pixel rechtsonder).
 - Projectie 4: over de zuidwestelijke diagonaal (pixel linksonder \rightarrow pixel rechtsboven).

We noemen dit *projectiemethode 2*. Bereken de vier bijbehorende projectiewaarden.

- Voer nu dezelfde stappen uit als bij opgave b, maar nu uitgaande van de projecties verkregen via projectiemethode 2.
- Zal het algoritme nu weer termineren voor $k = 2$?
- Zal het algoritme de correcte oplossing berekenen? (Met "correct" bedoelen we dat de oplossing de correcte pixelwaarden willekeurig dicht nadert als we MAX_ITER voldoende groot en ϵ voldoende klein nemen.)
Hint: bekijk de nulruimte bij projectiemethode 2.

Opgave 4 Algemene begrippen (15 pt)

Kies 7 begrippen uit de volgende lijst en leg van elk gekozen begrip in 3-5 regels de betekenis uit.

- Het centrale dogma van de moleculaire biologie
- Feed-forward neurale netwerk
- Genomics
- Gen-ontologie
- Ab initio* eiwitstructuurvoorspelling
- Dendrogram
- Phylogenetische boom
- Gen-expressie
- In silico* model
- Eiwit-DNA interacties
- Moleculaire switch
- Microarray
- Substitutiematrix
- Sequence logo