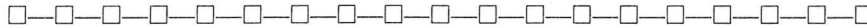


TENTAMEN INTRODUCTION TO BIOMEDICAL COMPUTING

17-4-2012



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 opgaven is het maximale aantal te behalen punten vermeld. Antwoorden dienen altijd van een motivatie te worden voorzien. Succes!

Opgave 1 Sequence alignment (25 pt)

Bekijk de volgende twee DNA sequenties: $s = \text{CTGA}$ en $t = \text{GCTG}$.

- (3 pt) Geef een korte beschrijving van het Needleman-Wunsch algoritme voor de alignment van de twee DNA sequenties s en t .
- (3 pt) Teken de dotplot van deze twee sequenties. Gebruik letters als symbool in de cellen van de dotplot.
- (12 pt) Bekijk nu het Needleman-Wunsch algoritme voor de recursieve berekening van de edit distance matrix D van de sequenties s en t . 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).

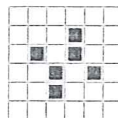
- (2 pt) Teken alle *trace-back* paden in de matrix D , beginnend rechtsonder en eindigend linksboven.
- (5 pt) Geef de optimale alignment(s) en bijbehorende edit distance(s).

Opgave 2 Cellulaire automaten (25 pt)

In deze opgave beschouwen we cellulaire automaten op een eindig tweedimensionaal rechthoekig rooster van cellen, waarbij elke cel 8 buurcellen heeft. Elke cel heeft twee mogelijke toestanden, levend of dood. Als overgangsregels ("transition rules") gebruiken we de *majority voting* (MV-regel) en de *Game of life* (GoL-regel).



(a)



(b)

- (5 pt) Geef een korte uitleg van de MV-regel en de GoL-regel.
- (15 pt) Schets, voor de twee beginconfiguraties (a) en (b) in bovenstaand plaatje, de configuratie in de volgende twee generaties, zowel bij toepassing van de MV-regel als de GoL-regel.
N.B. Levende cellen zijn zwart, dode cellen wit getekend.
- (5 pt) Beschrijf (zonder plaatjes te schetsen) het gedrag voor generaties hoger dan 2, weer voor beide configuraties (a) en (b), zowel bij toepassing van de MV-regel als de GoL-regel.

(Ga verder op de volgende pagina)

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  $\text{MAX\_ITER} \geq 1, \varepsilon > 0$ 
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  $\delta(\vec{f}^{(M)}, \vec{f}^{(0)}) < \varepsilon$  then
10:    break
11:  end if
12:   $\vec{f}^{(0)} \leftarrow \vec{f}^{(M)}$ 
13: end for

```

We bekijken de reconstructie van een 2×2 -plaatje uit vier projecties, de eerste twee over de rijen en de volgende twee over de kolommen. De projectiewaarden zijn weergegeven in het volgende plaatje:

6	f_1	f_2
14	f_3	f_4
	8	12

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

- a. (5 pt) Documenteer de body in de regels 4-13 van het Kaczmarz algoritme, d.w.z. beschrijf voor elke regel en/of bijeen horende reeks van regels welke operatie wordt uitgevoerd en met welk doel.
- b. (10 pt) Pas 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 $i = 1, 2, \dots, 4$. Initialiseer alle pixelwaarden op 0.
- c. (5 pt) De oplossing zal in de volgende iteratie niet meer veranderen, en het algoritme termineert voor $k = 2$. Leg uit waarom.
- d. (5 pt) Geef een voorbeeld van een andere oplossing (d.w.z., een verzameling waarden f_1, \dots, f_4 die verschillend zijn van de gevonden waarden in **b.**, maar die dezelfde projectiewaarden hebben als in bovenstaand plaatje). Verklaar waarom er meerdere oplossingen mogelijk zijn.

Opgave 4 Algemene begrippen (15 pt)

Kies 6 begrippen uit de volgende lijst en leg van elk gekozen begrip in 3-5 regels de betekenis uit.

1. Het centrale dogma van de moleculaire biologie
2. Gen-ontologie
3. Dendrogram
4. Hamming distance
5. Decision tree
6. Phylogenetische boom
7. Gen-expressie
8. Eiwit-DNA interacties
9. Supervised learning
10. Microarray
11. Substitutiematrix
12. Sequence logo