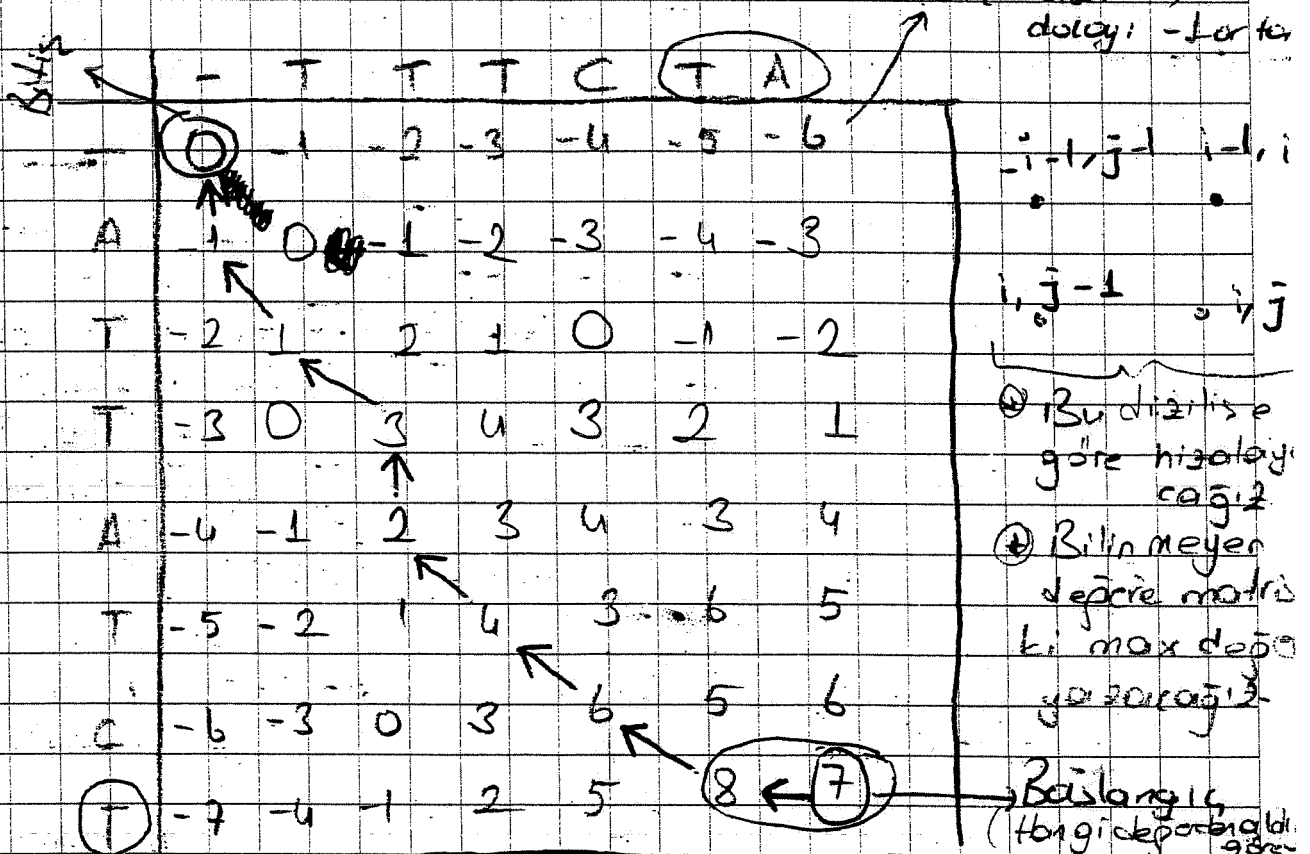


GLOBAL HİZALAMA (Dinamik Programlama) \* ÇIKARAK

Needleman and Wunsch Algoritması:

$$T(i, j) = \max \begin{cases} T(i-1, j-1) + \begin{matrix} \text{(match)} & \text{(mismatch)} \\ \text{Uyumu / Uyumsuzluğu} \end{matrix} \\ T(i-1, j) + \text{boşluk cezası} \\ T(i, j-1) + \text{boşluk cezası} \end{cases}$$

ATTATCT 3 Hizalamasını yapınız  
TTTCTA



\* (Uyumluluk = +2)  
Uyumsuzluk = 0  
Boşluk cezası = -1

0	-2	-1	-3	-2	-4	-3	-5	-3	-7
-2	1	-1	-1	-2	-2	-3	-3	-5	-3

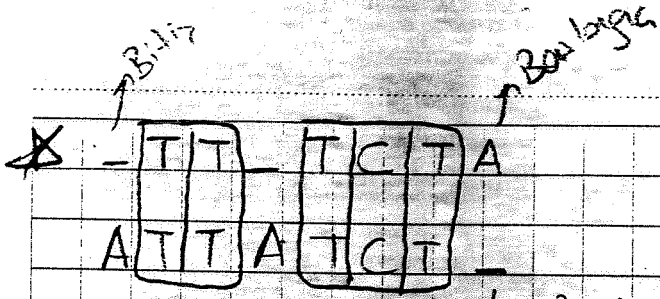
matrisdeki max değeri

matrisdeki max değeri

Uyumluluk

GIPTA

(2)



→ Bölüne T geldi.

$$10 - 3 = (7)$$

\* Dikeyler boşluk

\* Yataylar dolu

### LOKAL HIZALAMA (Smith ve Waterman):

Q K E S G P S S

K Q E S G L U R

\* Globalde mümkün  
olduğınca fazla  
hizalama işlemi.

\* Lokalde olabilirince  
+ dimal, yani yeni  
yeni karakterler  
fazla almalı.

\* Uyumlu = +2

Uyumsuz = -1

Boşluk = -2

\* Yönden aynı

= G G C T C A A T C A

-	0	0	0	0	0	0	0	0	0	0	0	Sevda L
A	0	0	0	0	0	0	2	2	0	0	2	en büyük
C	0	0	0	2	0	2	0	1	1	2	0	değerden
C	0	0	0	2	1	2	1	0	0	2	0	bulunuz
T	0	0	0	0	4	2	1	0	2	3	1	
A	0	0	0	0	2	3	4	3	1	1	3	CTCAA
A	0	0	0	0	0	1	5	6	4	2	3	CT-AA
G	0	2	2	0	0	0	3	4	5	3	1	* Sifiri
G	0	2	4	2	0	0	1	2	3	4	2	görene
												kardor oku
												ilerlet

3

27 03 2015

## EVOLÜSYONEL ALGORİTMALAR

### Genetik Algoritmalar

$$f(x,y) = x^3 - 11xy^2 + y^5 \quad x \geq 0 \quad y \geq 0 \quad x+y \leq 12 \text{ için}$$

$\max f(x,y)$ 'yi bulalım

### Önemli Aşamalar

- 1 - Kodlama  $\begin{cases} \text{İkili Kodlama} \\ \text{Gerçek değer Kodlama} \\ \text{Grey Kodlama} \end{cases}$
- 2 - Uygunluk değerinin bulunması (Fitness)  
Değerlendirme fonksiyonu

	x	y	GENLER	Değerlendirme fonk.
Birey (Kromozom)	0 0 1 0	0 0 0 1		50
	0 1 0 0	0 1 0 1		28
	1 0 0 0	0 0 0 1		10
				30
				15
100 birey				

①

X %5'lik (vs) en küçük dilimleri olup yerlerine, en büyük kullanıma belirliyor  
yük değerlerden kopyalama yaparız

- ② her zaman (toplami) ① den iyidir ④ Genetiklik  $\Rightarrow$  Caprazlama mutasyonu
- Ama ② ile ③ arasında böyle bir ilişki yoktur.

50
78
30
80
78
:

2

②


3

③

3  $\rightarrow$  1  
1. Jenerasyon  
2. "  
3. "

\* %100 ise  
%20 ise  
%80 ise  
%20 ise  
%80 ise

Genetik kaputite

\* Genetik oran %100  
ise bütün  
bireylerdir

1. Caprazlama (Herhangi iki birey)

$\rightarrow$  rastgele belirttik (caprazlama noktası)

01101000 } 01110111  
00010111 } 00010000 > Farklı iki birey elde ettik

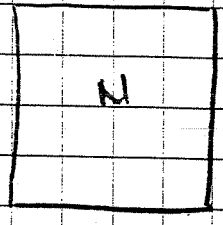
2. Mutasyon

00①10001 } Rastgele biri  
00⑧10001 } seçilerek 0  
değeri değiştirildi

Açıklama:  
Arka arkaya  
5 jenerasyon  
1340 bitlik  
0 noktası da  
kaybolur

5

# Knapsack Problem (Sırt Çantası Problemi)



$V_1, V_2, V_3, \dots, V_k$   
 $U_1, U_2, U_3, \dots, U_k$

⇒

1	2	3	...	k
1	0	1	...	0

\*  $\sum_{i: \text{aldığımız nesneler}} V_i \leq N$

1 → aldığımız

0 → almadığımız

\* Fitness = max  $\sum U_i$

→ aldığımız nesneler

KLAYTAN

\* Cluster Analysis

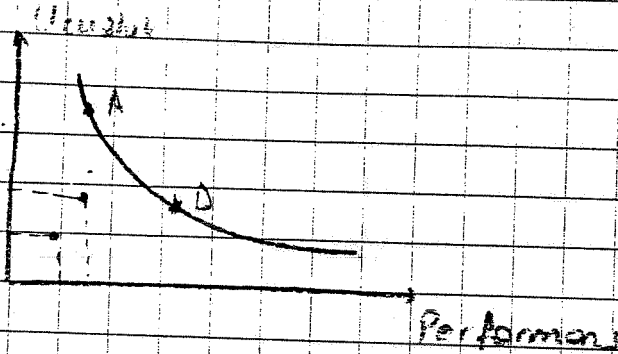
Clusterleme Yöntemi (Clustal W yöntemi)

↳ Ağırlıkları belirtmek için

6

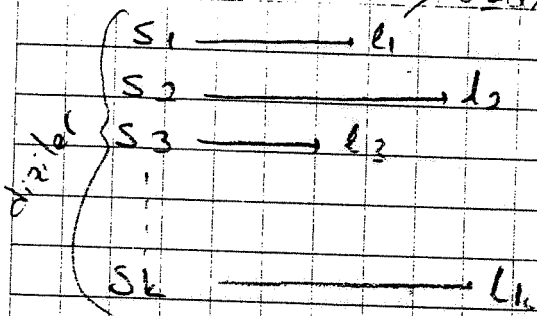
03.04.2015

multi-objektive optimizasyonu

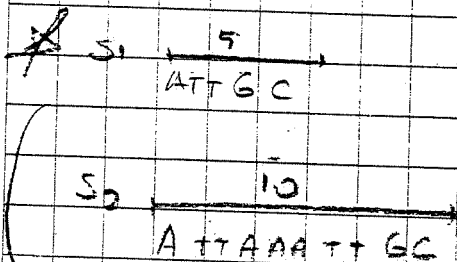


$f = w_1 \text{ performans} - w_2 \text{ uzunluk}$  // Amaç, maksimize edilmesi  
 ↓  
 ağırlık

Örnekli Dizi Hizalama (Genetik Algoritmalar ile)



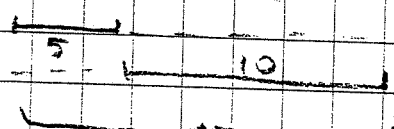
$A = (a_{ij})$   
 (Hizalama Durum)



uzunluklarında  
 min uzunluk 10 olur.  
 max değeri 15 olur.

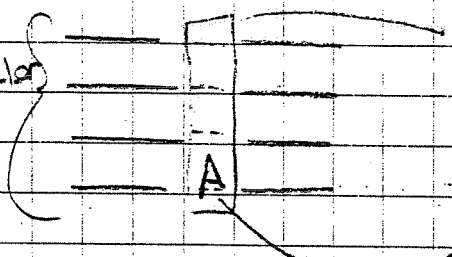
$A = \begin{pmatrix} ATT & GC \\ ATTAAATTGC \end{pmatrix}$

10 olur.

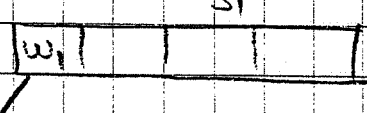


15 olur. GIPTA

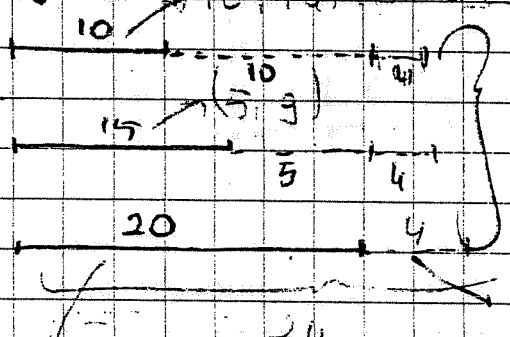
(7)

\* sequence (s)  Hizalamada bu konum tamamen bu dener  
 → Bir satırın dolu olması bu durumu göstermez koldur

\* Verilen sequence lerden bazılarının alınması daha iyi bir hizalama durumu olabilir

$S_1$   
  
 sequence  
 alınıp alınmayacağını belirler

\*  $G, l \rightarrow$  Sequence de kaç boşluk olacağını belirler

$\downarrow$   $\begin{matrix} \min & \max \\ 10 & 14 \end{matrix}$  (Boşluk sayıları)  
 \*  Hızalama yapıldıktan sonra max uzunluğun %20'sini geçmediği görülmüştür  
 %20'lik kum

$(10, 4)$   
 $\downarrow$   
 Hiç boşluk alınmayabilir

8

51

$w_1$	$ G $	$g_{12}$	$g_{13}$	---
-------	-------	----------	----------	-----

1. derece  $g_{11} = 3 \rightarrow ATA\_TTT\_CCC$   
 1. baslığı  $g_{12} = 4 \rightarrow ATA\_T\_$   
 $g_{13} = 3 \rightarrow ATA\_T\_$

1. baslık ten önce  
 3 tane sütunluk d  
 alınmalı

AACT  
 CACT  
 TTT  
 TAAT  
 ACT

	1	2	3	4
A	0,25	1	—	—
C	0,25	0	0,8	—
T	0,5	0	0,2	1
G	0	0	—	—

Bu değerler  
 sütunlardaki  
 bulunma  
 oranları

Her sütunun en  
 büyük değeri alınır

$$\frac{0,5 + 1 + 0,8 + 1}{4} = 0,825 \quad // \quad \text{Eğer } \frac{4}{4} = 1 \text{ olursa}$$

Baslık kaldırılmış  
 gösterim

A — — —  
 — T — —  
 — — C —  
 — — — T



9

\* Hizalamada sequenceler arasındaki benzerlikleri bulunuz.

# A A A G A A A T T C A  
A A A A T C A  
5 5 5 5

↳ Cezası  $5 + 5 + 5 + 5 = 20$

A A A G A A A T T C A  
A A A - - - T T C A  
5 2 2 2

↳ Cezası  $5 + 2 + 2 + 2 = 11$

Bosluklar  
ard arda  
geldiğinde  
cezalıdır.

\* İyi hizalama boşluğun az ve benzerliğin fazla  
olduğu hizalamadır.

\* Ama bütün sequencelerin kullanılması ama daha  
iyi durumlar için bazı sequenceler devre dışı  
bırakılabilir.

(10)

10 04 2015

# FMGA Algorithms, #

Motif: TATA box motif = [TATAAA]

BRE motif = [(G/C)(G/A)CGCC]

Tat motif = [TCA(G/T)T(T/C)]

DPE motif = [(A/G)G(A/T)(C/T)(G/A/C)]

\* P: ACCGTA

ACCGTA
.
.
.
.
.

\* Aranan motif verilen sekanslar  
birbirine karşılaştırmalı olarak  
maksimum olan ol (Eithenman)

NOT: Birbirine karşılaştırmalı olarak

m: A/C

r: A/G

w: A/T

s: C/G

y: C/T

k: G/T

P: MCCGTA

S: TCCGTA TACCGTA ACCGTA ACCGTA

sequence

5/6

5/6

5/6

} Sırtla

Verilen  
değerler  
göre hesap  
lanır

Result: 7 → motif uzerinde

TFS

7x6=42 → 30 tane sekans  
bulunur

TFS(%)

(12/42) = 1

7 tane  
sekans  
bulunur

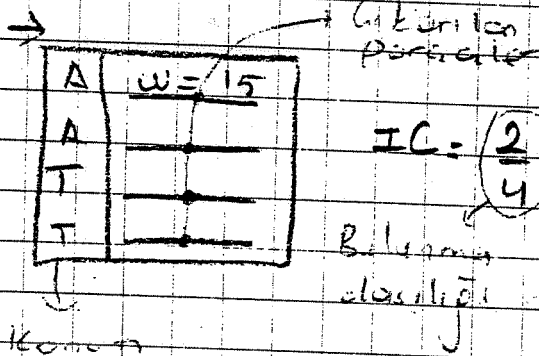
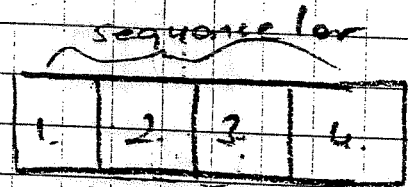
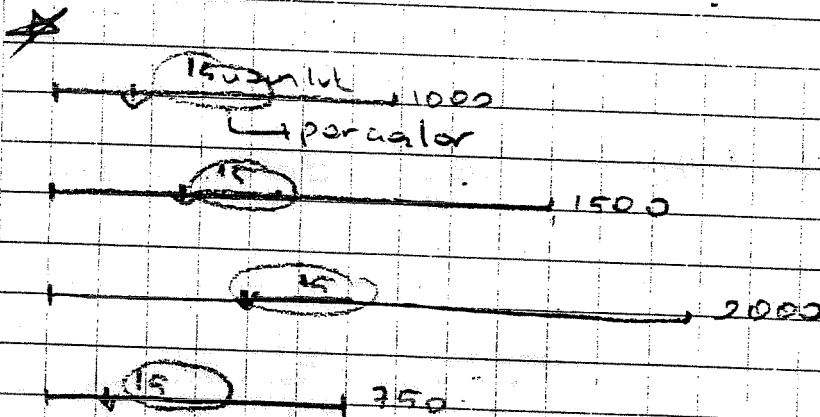
Herkesin Hayatına Kimse Karışmaz!

(11)

24 04 2015

Cuma

Motif bulma (Disardan motif uzunluğu belirlenir)



$$IC = \left( \frac{2}{4} \right) \log_2 \frac{2}{4} + 2 \log_2 \frac{2}{4}$$

0,25 → 1/4

$$\begin{matrix} A \\ T \\ G \\ C \end{matrix} \left. \vphantom{\begin{matrix} A \\ T \\ G \\ C \end{matrix}} \right\} \text{ tüm } IC = \frac{1}{4} \log_2 \frac{0,25}{0,25} + 0 + 0 + 0$$

↓ ↓ ↓  
A T G C

$$= \underline{\underline{0}}$$

$$IC = \sum i.b \log_2 \frac{f.b}{p.b}$$

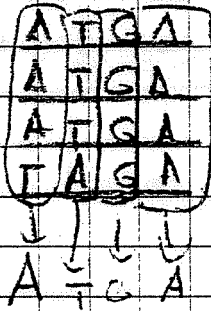
Fitness

$$Fitness = \sum_{i=1}^w IC$$

motif uzunluğu  
Motif uzunluğunun disardan verilmesi

## Consensus Reaktif

Δ Matriflerde en fazla bulunan bölümleri değeri olarak kabul edilir.



⊗ Bütün motif uzunluğunun hepsi aynı harflerden oluşmuştur.