1) Her canlı ayrı küme

① Kümeleme

② Mesafe matrisi : a=1 olunca Minkowski metriği Manhattan yöntemi ile uzaklık ölçer.

$$d_{ij} = \sum_{k=1}^{p} |x_{ik} - x_{jk}|$$

$d_{AB} = |3-5| + |6-4| + |4-10| = 10$

$d_{AC} = |3-8| + |6-3| + |4-9| = 13$

$d_{AD} = |3-10| + |6-5| + |4-3| = 9$

$d_{AE} = |3-3| + |6-5| + |4-3| = 2$

$d_{BC} = |5-8| + |4-3| + |10-9| = 4$

$d_{BD} = |5-10| + |4-5| + |10-3| = 13$

$d_{BE} = |5-3| + |4-5| + |10-3| = 10$

$d_{CD} = |8-3| + |3-5| + |9-3| = 13$

$d_{CE} = |8-3| + |3-5| + |9-3| = 13$

$d_{DE} = |10-3| + |5-5| + |3-3| = 7$

OTU	A	B	C	D	E
A	0				
B	10	0			
C	13	4	0		
D	9	13	13	0	
E	2	10	13	7	0

En yakın kümeleri birleştirdim.

"group average linkage":

$$d_{ij} = \frac{\Sigma_i \Sigma_j}{N_i N_j} = \frac{d_{AE}}{\text{özellik sayısı}} = \frac{2}{3} = 0.66$$

OTU	(AE)	B	C	D
(AE)	0			
B	10	0		
C	13	4	0	
D	8	13	13	0

$d_{(AE)-B} = |3-5| + |5.5-4| + |3.5-10| = 10$

$d_{(AE)-C} = |3-8| + |5.5-3| + |3.5-9| = 13$

$d_{(AE)-D} = |3-10| + |5.5-5| + |3.5-3| = 8$

Şimdi bulunan matriste ise B ile C yakın. Onları kümeliyorum.

(A E) (B C) . D

OTU	(AE)	(BC)	D
(AE)	0		
(BC)	12	0	
D	8	11	0

$d_{(AE)-(BC)} = |3-7| + |5.5-3,5| + |3.5-9.5| = 12$

$d_{(AE)-D} = 8$

$d_{(BC)-D} = |7-10| + |3.5-5| + |9.5-3| = 11$

A ile D en yakın çıktı

Son kümeleme

(A E D) (B C)

2) $P(A) = 0,15$

$P(C) = 0,40$

$P(G) = 0,35$

$P(T) = 0,10$

	-4	-3	-2	-1	+1	+2	+3	+4
sekans 1	T	C	G	C	C	G	T	C
sekans 2	C	A	C	A	A	G	G	G
sekans 3	G	G	T	C	A	C	G	T
sekans 4	T	G	C	C	G	A	G	T
sekans 5	G	G	C	C	G	C	A	C
sekans 6	C	G	T	A	T	C	C	T
sekans 7	G	A	A	C	G	T	G	C
sekans 8	C	C	T	A	A	T	G	T
sekans 9	T	C	A	T	T	G	C	T
sekans 10	C	A	G	A	T	A	T	T

Entropi:

$$H(x) = -\sum_{j=1}^{J} P(x_j) \log_2 P(x_j)$$

-4. baz için $\Rightarrow$ $H(x) = -\left[\underbrace{(0,10 \times \log_2(0,10)) \times 3}_{\text{T'lerin ihtimali}} + \underbrace{(0,35 \times \log_2(0,35) \times 3}_{\text{G ihtimali}} + (0,40 \times \log_2(0,40) \times 4) \right] = 4,7$

-3. baz için $\Rightarrow$ $H(x) = -\left[(0,15 \times \log_2(0,15) \times 3) + (0,4 \times \log_2(0,4) \times 3) + (0,35 \times \log_2(0,35) \times 4) \right] = 4,94$

-2. baz için $\Rightarrow$ $H(x) = -\left[(0,15 \times \log_2(0,15) \times 2) + (0,40 \times \log_2(0,4) \times 3) + (0,35 \times \log_2(0,35) \times 2) + (0,1 \times \log_2(0,1) \times 3) \right] = 4,46$

-1. baz için $\Rightarrow$ $H(x) = -\left[(0,15 \times \log_2(0,15) \times 4) + (0,4 \times \log_2(0,4) \times 5) + (0,1 \times \log_2(0,1) \times 1) \right] = 4,62$

1. baz için $\Rightarrow$ $H(x) = -\left[(0,15 \times \log_2(0,15) \times 3) + (0,4 \times \log_2(0,4) \times 1) + (0,35 \times \log_2(0,35) \times 3) + (0,1 \times \log_2(0,1) \times 3) \right] = 4,35$

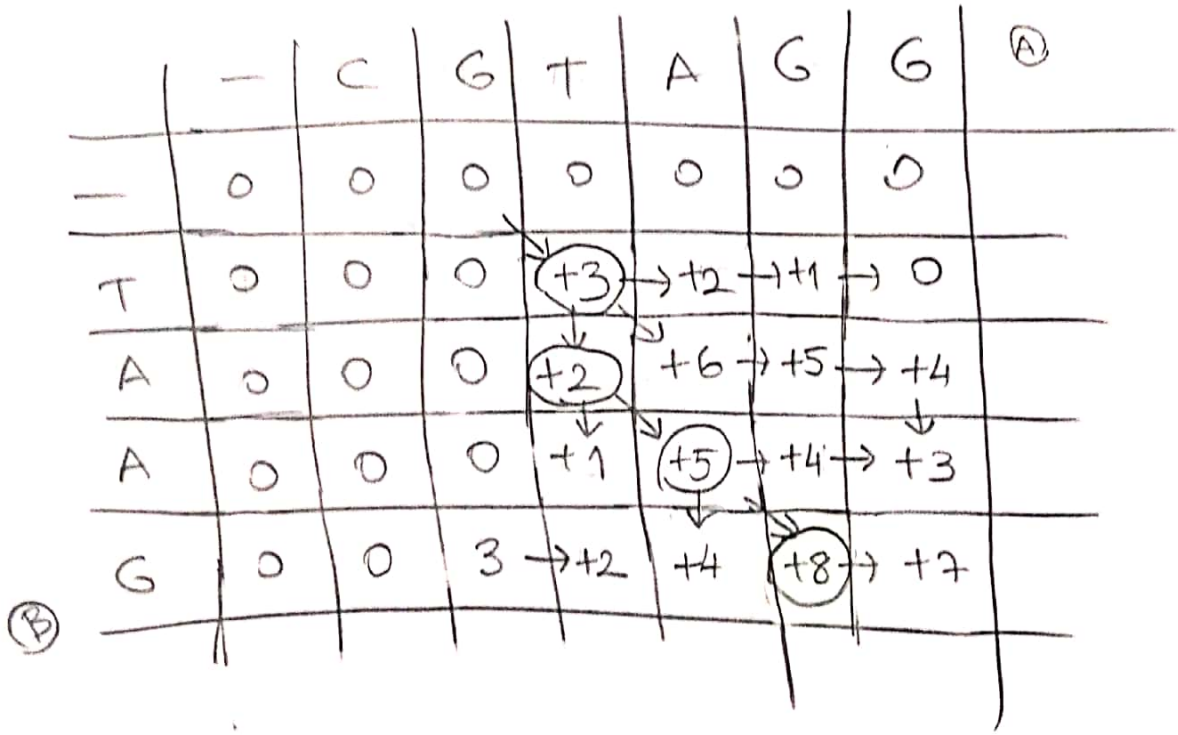
2. baz için $\Rightarrow$ $H(x) = -\left[(0,15 \times \log_2(0,15) \times 2) + (0,4 \times \log_2(0,4) \times 3) + (0,35 \times \log_2(0,35) \times 2) + (0,1 \times \log_2(0,1) \times 2) \right] = 4,66$

3. baz için $\Rightarrow H(X) = -\left[(0{,}15 \times \log_2(0{,}15) \times 1) + (0{,}4 \times \log_2(0{,}4) \times 2) + (0{,}35 \times \log_2(0{,}35) \times 5) + (0{,}1 \times \log_2(0{,}1) \times 2\right] = 4{,}78$

4. baz için $\Rightarrow H(X) = -\left[(0{,}4 \times \log_2(0{,}4) \times 3 + (0{,}35 \times \log_2(0{,}35) \times 1) + (0{,}1 \times \log_2(0{,}1) \times 6\right] = 4{,}11$

3) match → +3 ↘ gap open / extend → -1

mismatch → -2



A : T _ A G

B : T A A G

Skor : 8