

Exercise #4

We have given S3 and S4 sessions and we need to determine the best (optimal) alignment by Needleman-Wunsch Algorithm.

S3: <P1, P6, P1, P4>

S4: <P5, P4, P1, P6, P7>

Given,

$S(P4, P6) = 0.9$, $S(P5, P7) = 0.9$

$S(P_i, P_j) = 1$ if $i=j$ and $S(P_i, P_j)=0$ if $i \neq j$

Page Similarity

Combination of P1 -> {P5, P4, P1, P6, P7}

$S(P1, P5) = 0$

$S(P1, P4) = 0$

$S(P1, P1) = 1$ ($i=j$)

$S(P1, P6) = 0$

$S(P1, P7) = 0$

Combination of P6 -> {P5, P4, P1, P6, P7}

$S(P6, P5) = 0$

$S(P6, P4) = 0$

$S(P6, P1) = 0$

$S(P6, P6) = 1$ ($i=j$)

$S(P6, P7) = 0$

Combination of P4 -> {P5, P4, P1, P6, P7}

$S(P4, P5) = 0$

$S(P4, P4) = 1$ ($i=j$)

$S(P4, P1) = 0$

$S(P4, P6) = 0.9$ (Given)

$S(P4, P7) = 0$

I am going to use score as given in paper,

```
//a pair of Web pages with similarity 1  
if matching then score = 20;
```

```
//a pair of Web pages with similarity 0  
else if mis-matching then score = -10;
```

```
//a Web page aligns with a gap  
else if gap then score = -10;
```

```
//the pair of Web pages with similarity between 0 and 1  
else score = -10 ~ 20;
```

$S(X_i, Y_j) = (-10 + 30 \times \text{Page_Similarity})$

I am taking gap as $d = -10$

$$A(i, j) = \text{MAX}[\{A(i-1, j-1) + S(X_i, Y_j)\}; A(i-1, j)+d; A(i, j-1)+d]$$

- $A(P1, P5) = \text{MAX}[(0 + -10); (-10 + -10); (-10 + -10)] = [-10; -20; -20] = -10$
 $A(P1, P4) = \text{MAX}[(-10 + -10); (-20 + -10); (-10 + -10)] = [-20; -30; -20] = -20$
 $A(P1, P1) = \text{MAX}[(-20 + 20); (-30 + -10); (-20 + -10)] = [0; -40; -30] = 0$
 $A(P1, P6) = \text{MAX}[(-30 + -10); (-40 + -10); (0 + -10)] = [-40; -50; -10] = -10$
 $A(P1, P7) = \text{MAX}[(-40 + -10); (-50 + -10); (-10 + -10)] = [-50; -60; -20] = -20$

 $A(P6, P5) = \text{MAX}[(-10 + -10); (-10 + -10); (-20 + -10)] = [-20; -20; -30] = -20$
 $A(P1, P5) = \text{MAX}[(-20 + -10); (-20 + -10); (-30 + -10)] = [-30; -30; -40] = -30$
 $A(P4, P5) = \text{MAX}[(-30 + -10); (-30 + -10); (-40 + -10)] = [-40; -40; -50] = -40$

 $A(P6, P4) = \text{MAX}[(-10 + -10); (-20 + -10); (-20 + -10)] = [-20; -30; -30] = -20$
 $A(P6, P1) = \text{MAX}[(-20 + -10); (0 + -10); (-20 + -10)] = [-30; -10; -30] = -10$
 $A(P6, P6) = \text{MAX}[(0 + 20); (-10 + -10); (-10 + -10)] = [20; -20; -20] = 20$
 $A(P6, P7) = \text{MAX}[(-10 + -10); (-20 + -10); (20 + -10)] = [-20; -30; 10] = 10$

 $A(P1, P4) = \text{MAX}[(-20 + -10); (-20 + -10); (-30 + -10)] = [-30; -30; -40] = -30$
 $A(P4, P4) = \text{MAX}[(-30 + 20); (-30 + -10); (-40 + -10)] = [-10; -40; -50] = -10$

 $A(P1, P1) = \text{MAX}[(-20 + 20); (-10 + -10); (-30 + -10)] = [0; -20; -40] = 0$
 $A(P1, P6) = \text{MAX}[(-10 + -10); (20 + -10); (0 + -10)] = [0; 10; -10] = 10$
 $A(P1, P7) = \text{MAX}[(20 + -10); (10 + -10); (10 + -10)] = [10; 0; 0] = 10$

 $A(P4, P1) = \text{MAX}[(-30 + -10); (0 + -10); (-10 + -10)] = [-40; -10; -20] = -10$

 $A(P4, P6) = \text{MAX}[(0 + 17); (10 + -10); (-10 + -10)] = [17; 0; -20] = 17$
 $A(P4, P7) = \text{MAX}[(10 + 0); (10 + -10); (17 + -10)] = [10; 0; 7] = 10$

	-	P5	P4	P1	P6	P7
-	0	-10	-20	-30	-40	-50
P1	-10	-10	-20	0	-10	-20
P6	-20	-20	-20	-10	20	10
P1	-30	-30	-30	0	10	10
P4	-40	-40	-10	-10	17	10

Best (Optimal) alignment score is 20

Session similarity = $20/5 = 4$