## **HOMEWORK 05**

#### INTRODUCTION:

One method of solving problems is to divide them into smaller subproblems and solve each subproblem only once, a process known as Dynamic programming. When the problem can be broken down into overlapping subproblems and the answers to these subproblems are saved and utilized again to cut down on computation that would otherwise be necessary, it is especially helpful.

There are two primary categories of dynamic programming: The top-down approach and the Bottom-up approach.

- In Top-down dynamic programming, sometimes referred to as memoization, solves larger subproblems recursively after completing the main problem. To prevent duplicating computation, the solutions are stored in a table.
- In Bottom-up dynamic programming, the smallest subproblems are solved first, and the larger subproblems are solved iteratively using the solutions to the smaller subproblems until the original problem is solved.

For local sequence alignment, the algorithm known as Smith-Waterman is a dynamic programming technique. Even in the case of gaps and mismatches, it is especially helpful for locating similar regions between two biological patterns such as DNA, RNA, or protein sequences.

It works in the following ways:

- a. Initialization: Make a matrix in which the rows and columns represent the characters in the two sequences that are being compared. This matrix is frequently referred to as the score/similarity matrix. Put zeros in the first row and column and start.
- b. Scoring methods: Assign points for gaps, mismatches, and matches. A match usually receives a positive score, a mismatch receives a negative score, and opening and extending a gap results in penalties.
- c. Completing the matrix: Work through the matrix, assigning a score to each cell based on the scoring scheme and the adjacent cells. Each cell's score is the highest possible score for all alignments that could end at that particular position.
- d. Traceback: Upon filling the matrix, identify the cell that has the highest score. The alignment ends with this cell. From this cell, trace backward along the path of the highest scores until it comes to a cell that has a score of zero, designating the alignment's beginning. The ideal local alignment is obtained from this traceback.
- 1.1 Bottom-Up Approach for SmithWaterman Algorithm:

```
void SW_bottomUp(char* X, char* Y, char** P, int** H, int n, int m) {
    for (int i = 0; i <= n; ++i) {
        for (int j = 0; j <= m; ++j) {
            H[i][j] = 0;
            P[i][j] = '-';
        }
}

for (int i = 1; i <= n; ++i) {
        int match = H[i - 1][j - 1] + (X[i - 1] == Y[j - 1] ? 2 : -1);
        int gapX = H[i - 1][j] - 1;
        int gapY = H[i][j] - 1] - 1;

        if (match >= gapX && match >= gapY && match >= 0) {
            H[i][j] = match;
            P[i][j] = '\\';
        } else if (gapX >= gapY && gapX >= 0) {
            H[i][j] = gapX;
            P[i][j] = '-';
        } else if (gapY >= 0) {
            H[i][j] = gapY;
            P[i][j] = '-';
        } else {
            H[i][j] = 0;
            P[i][j] = '-';
        }
}
```

1.2 Top-Down Approach for Smith-Waterman Algorithm:

1.3 Printing PRINT-SEQ-ALIGN-X & PRINT-SEQ-ALIGN-Y

#### 1.4 Testing the Algorithm

The H and P are initilzed to zero completely. Then looking up from the previous values on the table the directions and costs are put back to the table. The conditions used are

$$c[i,j] = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0, \\ c[i-1,j-1]+1 & \text{if } i,j > 0 \text{ and } x_i = y_j, \\ \max(c[i,j-1],c[i-1,j]) & \text{if } i,j > 0 \text{ and } x_i \neq y_j. \end{cases}$$

Result for Smith Waterman for:

X= dcdcbacbbb

Y = acdccabdbb

	acaccac	Juou									
Н		a	c	d	c	c	a	b	d	b	b
	0	0	0	0	0	0	0	0	0	0	0
d	0	0	0	2	1	0	0	0	2	1	0
c	0	0	2	1	4	3	2	1	1	1	0
d	0	0	1	4	3	3	2	1	3	2	1
С	0	0	2	3	6	5	4	3	2	2	1
b	0	0	1	2	5	5	4	6	5	4	4
a	0	2	1	1	4	4	7	6	5	4	3
С	0	1	4	3	3	6	6	6	5	4	3
b	0	0	3	3	2	5	5	8	7	7	6
b	0	0	2	2	2	4	4	7	7	9	9
b	0	0	1	1	1	3	3	6	6	9	11

P	0	a	c	d	c	c	a	b	d	b	b
0	0	0	0	0	0	0	0	0	0	0	0
d	0	d	d	d	1	d	d	d	d	1	1
c	0	d	d	1	d	d	1	1	u	d	d
d	0	d	u	d	1	1	d	d	d	1	1
c	0	d	d	u	d	d	1	1	1	d	d
b	0	d	u	u	u	d	d	d	1	d	d
a	0	d	1	u	u	d	d	1	d	d	d
c	0	u	d	1	d	d	u	d	d	d	d
b	0	u	u	d	d	u	d	d	1	d	d
b	0	d	u	d	d	u	d	d	d	d	d
b	0	d	u	d	d	u	d	d	d	d	d

The values of X= dcdcbacbbb

The values of Y = acdccabdbb

The values of X'= cdcbacb\_bb

The values of Y'= cdcca\_bdbb

The maximum alignment for X and Y are solved and X' and Y' are obtained. The maximum cost is 11.

# **2. OPTIMAL BINARY SEARCH TREE:**

An optimal binary search tree (BST) is a data structure that aims to reduce the mean length of search for a given set of keys, each with its probability of being searched. Dynamic programming is used in the setting up of an optimal BST, where the lowest expected search cost of building subtrees is calculated using a cost matrix. The average search time is lowered by placing keys with higher probability closer to the root. Optimal BSTs offer a balance between search time and storage capacity.

Given problem states,

i	1	2	3	4	5
p i	0.21	0.15	0.28	0.12	0.24

→ The matrix, w,e, and R are initialized accordingly. The w is the weight matrix, e is the Expected cost matrix and the R is the root matrix.

W	0	1	2	3	4	5
1	0					
2		0				
3			0			
4				0		
5					0	
6						0

e	0	1	2	3	4	5	
1	0						

2	0				
3		0			
4			0		
5				0	
6					0

R	1	2	3	4	5
1	1				
2		2			
3			3		
4				4	
5					5

The matrix then goes into a main for-loop that iterates for l works on the subtrees with l keys. There the order of the subtree size from smaller to larger is iterated and the values are calculated.

Calculations are shown below:

For 
$$i = 1$$
,  $j = 2$ :

$$\begin{split} &w[1,2] = w[1,1] + p_2 = 0.21 + 0.15 = 0.36 \\ &e[1,2] = e[1,0] + e[2,2] + w[1,2] = 0 + 0.15 + 0.36 = 0.51 \Rightarrow r = 1 \\ &e[1,2] = e[1,1] + e[3,2] + w[1,2] = 0.21 + 0 + 0.36 = 0.57 \Rightarrow r = 2 \end{split}$$

$$e[1, 2] = min(0.51, 0.57) = 0.51 \rightarrow R[1,2] = 1$$

For 
$$i = 2$$
,  $j = 3$ :

$$\begin{split} &w[2,3] = w[2,2] + p_3 = 0.15 \ + 0.28 = 0.43 \\ &e[2,3] = e[2,1] + e[3,3] + w[1,2] = 0 + 0.28 + 0.43 = 0.71 \Rightarrow r = 2 \\ &e[2,3] = e[2,2] + e[4,3] + w[1,2] = 0.15 + 0 + 0.43 = 0.58 \Rightarrow r = 3 \end{split}$$

$$e[2, 3] = min(0.71, 0.58) = 0.58 \rightarrow R[2, 3] = 3$$

For 
$$i = 3$$
,  $j = 4$ :

$$w[3, 4] = w[3, 3] + p_4 = 0.28 + 0.12 = 0.4$$

$$e[3, 4] = e[3, 2] + e[4,4] + w[3,4] = 0 + 0.12 + 0.4 = 0.52 \rightarrow r = 3$$

$$e[3, 4] = e[3, 3] + e[5, 4] + w[3, 4] = 0.28 + 0 + 0.4 = 0.68 \rightarrow r = 4$$

$$e[3, 4] = min(0.52, 0.68) = 0.52 \rightarrow R[3, 4] = 3$$

For 
$$i = 4$$
,  $j = 5$ :

$$w[4, 5] = w[4, 4] + p_5 = 0.12 + 0.24 = 0.36$$

$$e[4, 5] = e[4, 3] + e[5, 5] + w[4, 5] = 0 + 0.24 + 0.36 = 0.6 \rightarrow r = 4$$

$$e[4, 5] = e[4, 4] + e[6,5] + w[4,5] = 0.12 + 0 + 0.36 = 0.48 \rightarrow r = 5$$

$$e[4, 5] = min(0.6, 0.48) = 0.58 \rightarrow R[4,5] = 5$$

For 
$$i = 3$$
,  $j = 5$ :

$$w[3, 5] = w[3, 4] + p_5 = 0.4 + 0.24 = 0.64$$
  
 $e[3, 5] = e[3, 2] + e[4,5] + w[4,5] = 0 + 0.48 + 0.64 = 1.12 \rightarrow r = 3$ 

$$\begin{array}{l} e[3,5] = e[3,3] + e[5,5] + w[4,5] = 0.28 + 0.24 + 0.64 = 1.16 \Rightarrow r = 4\\ e[3,5] = e[3,4] + e[6,5] + w[4,5] = 0.58 + 0 + 0.64 = 1.16 \Rightarrow r = 5\\ e[3,5] = \min(1.12,1.16,1.16) = 1.12 \Rightarrow R[3,5] = 3\\ \end{array}$$

$$For i=2,j=4:$$

$$w[2,4] = w[2,3] + p_4 = 0.43 + 0.12 = 0.55\\ e[2,4] = e[2,1] + e[3,4] + w[2,4] = 0 + 0.52 + 0.55 = 1.07 \Rightarrow r = 2\\ e[2,4] = e[2,2] + e[4,4] + w[2,4] = 0.15 + 0.12 + 0.55 = 0.82 \Rightarrow r = 3\\ e[2,4] = e[2,3] + e[5,4] + w[2,4] = 0.58 + 0 + 0.55 = 1.13 \Rightarrow r = 4\\ e[2,4] = \min(1.07,0.82,1.13) = 0.82 \Rightarrow R[2,4] = 3\\ \end{array}$$

$$For i=1,j=3:$$

$$w[1,3] = w[1,2] + p_3 = 0.36 + 0.28 = 0.64\\ e[1,3] = e[1,0] + e[2,3] + w[1,3] = 0.21 + 0.28 + 0.64 = 1.22 \Rightarrow r = 1\\ e[1,3] = e[1,1] + e[3,3] + w[1,3] = 0.21 + 0.28 + 0.64 = 1.13 \Rightarrow r = 2\\ e[1,3] = e[1,2] + e[4,3] + w[1,3] = 0.51 + 0 + 0.64 = 1.15 \Rightarrow r = 3\\ e[1,3] = min(1.22,1.13,1.15) = 1.13 \Rightarrow R[1,3] = 2\\ \end{array}$$

$$For i=2,j=5:$$

$$w[2,5] = w[2,4] + p_5 = 0.55 + 0.24 = 0.79\\ e[2,5] = e[2,1] + e[3,5] + w[2,5] = 0 + 1.12 + 0.79 = 1.91 \Rightarrow r = 2\\ e[2,5] = e[2,2] + e[4,5] + w[2,5] = 0.15 + 0.48 + 0.79 = 1.61 \Rightarrow r = 3\\ e[2,5] = e[2,3] + e[5,5] + w[2,5] = 0.82 + 0 + 0.79 = 1.61 \Rightarrow r = 5\\ e[2,5] = e[2,4] + e[6,5] + w[2,5] = 0.82 + 0 + 0.79 = 1.61 \Rightarrow r = 5\\ e[2,5] = e[1,4] + e[6,5] + w[2,5] = 0.82 + 0 + 0.79 = 1.61 \Rightarrow r = 5\\ e[2,5] = e[1,4] + e[4,4] + w[1,4] = 0 + 0.82 + 0.76 = 1.58 \Rightarrow r = 1\\ e[1,4] = e[1,0] + e[2,4] + w[1,4] = 0.21 + 0.52 + 0.76 = 1.58 \Rightarrow r = 1\\ e[1,4] = e[1,1] + e[3,4] + w[1,4] = 0.21 + 0.52 + 0.76 = 1.49 \Rightarrow r = 2\\ e[1,4] = e[1,2] + e[4,4] + w[1,4] = 0.21 + 0.52 + 0.76 = 1.49 \Rightarrow r = 2\\ e[1,4] = e[1,3] + e[5,4] + w[1,4] = 0.21 + 0.52 + 0.76 = 1.49 \Rightarrow r = 3\\ e[1,4] = e[1,2] + e[4,4] + w[1,4] = 0.21 + 0.52 + 0.76 = 1.49 \Rightarrow r = 3\\ e[1,4] = e[1,2] + e[4,4] + w[1,4] = 0.21 + 0.52 + 0.76 = 1.49 \Rightarrow r = 3\\ e[1,4] = e[1,3] + e[5,4] + w[1,5] = 0.21 + 1.12 + 12.233 \Rightarrow r = 2\\ e[1,4] = e[1,2] + e[4,4] + w[1,4] = 0.21 + 0.24 \Rightarrow r = 1\\ e[1,5] = e[1,4] + e[4,5] + w[1,5] = 0.21 + 1.12 + 12.233 \Rightarrow r = 2\\ e[1,5] = e[1,4] + e[4,5] + w[1,5] = 0.21 + 1.42 + 1.0 = 2.42 \Rightarrow r = 1\\ e[1,5] = e[1,4] + e[6,5] + w[1,5] = 0.44 + 1 = 2.48$$

The final w, e and R matrix are shown below:

W	0	1	2	3	4	5
1	0	0.21	0.36	0.64	0.76	1.0
2		0	0.15	0.43	0.55	0.79
3			0	0.28	0.4	0.64
4				0	0.12	0.36
5					0	0.24
6						0

e	0	1	2	3	4	5
1	0	0.21	0.51	1.13	1.39	<mark>1.99</mark>
2		0	0.15	0.58	0.82	1.42
3			0	0.28	0.52	1.12
4				0	0.12	0.48
5					0	0.24
6						0

R	1	2	3	4	5
1	1	1	2	3	3
2		2	3	3	3
3			3	3	3
4				4	5
5					5

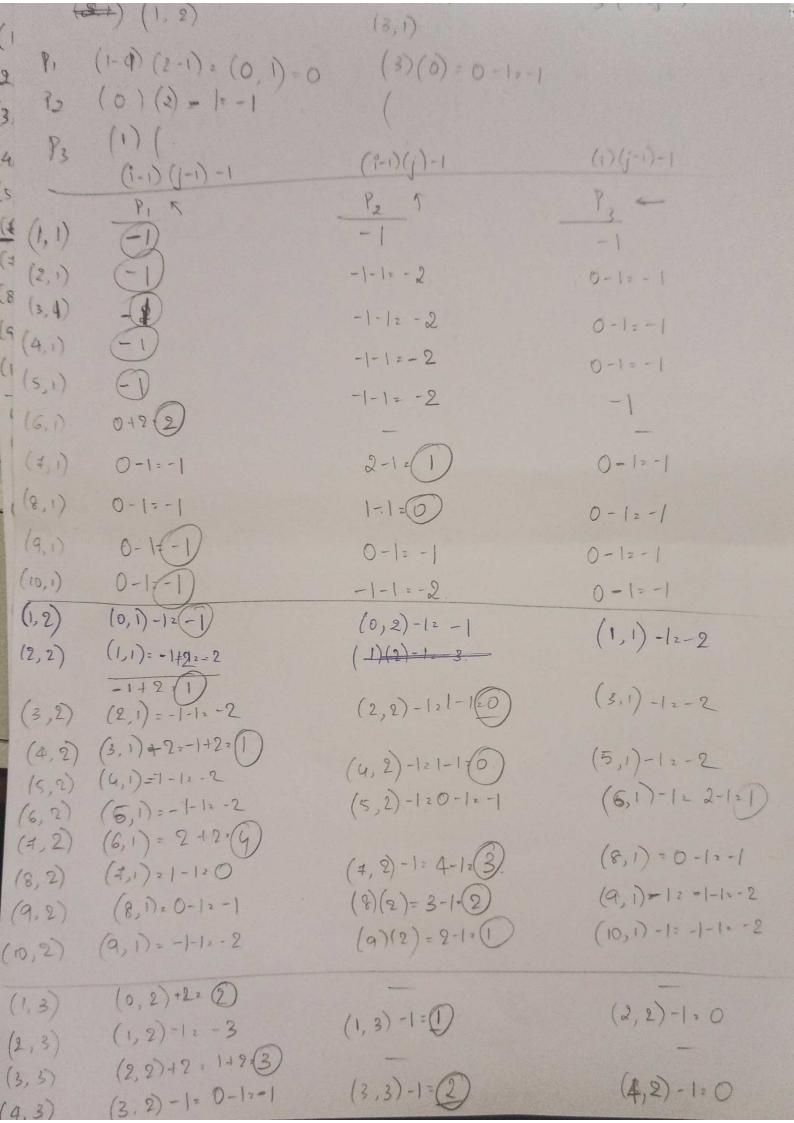
The cost of the given BST is 1.99.

The tree gets k3 as the root and the k2 as left child and k5 as the right child of the root. the k1 will be the left child of k2 and k4 will be the left child of k5.

### **REFERENCES:**

- [1] Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009) *Introduction to Algorithms* (3rd ed.). The MIT Press.
- [2] Sedgewick, R., & Wayne, K. (2011). Algorithms (4th Edition). Addison-Wesley Professional.

	0	و	90	44		\	9	0		W T	21		- T	1/1	
u× l	0-	9	12	.   "	-	10	0	U		2	0	a	1	/×	
98	0	0	1	5/1	0	0	0	0		0	. 0	0	0	1	0
100	0	-	-	9	1	-	0	1-	1	2	10	山	0.	2	-
, 88	0	+-		-	-	20	W	1	4	_	1-	115	0	0	2
00	-	7	0	w	-	12	1	1	2	6	0	1-15	6	2:	lo
0 8	-	18	20	20	W	14	5		Kr	80	-	1-3	0	0	2
	4	2	t	01	10	4	1-7	>	4		0	1	0.	0	ó
	1	2	>	21	19	10	).(	N	W	0	-	20	0	0	6
80	(	N	4	1-2	al	1/2	1	N	20	0	1	-	0	0	*
00		OI	0	16	14	4		1	80.	8	W	0	0	0-	ab
		00	00	6	6	, v	, (	N	2	20	Ro	1-	0	6	× × × × × × ×
		18	R	2	U	180	1	6	-	-	-	-	0	<u>o</u>	e de ce
					-									16	X= dcd c backbb



	P, (1-1)(j-1)	P2(1-1)(j)	P3 (1) (9-1)
(5,3)	(4,2)-1=0	(4,3)-1=2-1=1	(5,2)-1=-1
(6,3)	(5,2)-1=-1	(5,3)-1=0	(6,2)-1=0
(7,3)	(6,2)-1=0	(6,3)-1=0-1=-1	(7,2)-1=3
(8,3)	(7,2)-1=4-1=3	(7,3)-1=3-1=2	(8,2)-1=3-1=2
(9,3)	(8,2)-1:3-1:2	(8,3)-1=3-1=2	(9,2)-1=2-1,1
(10,3)	(9,2)-1=2-1=1	(9,3)-1=2-1=1	(10, 2) -1 20
(1,4)	(0,3)-1=-1	(0, u)-1=-1	(1,3)-122-120
(2,4)	(1,3)+2 2 (4)	6 -	
(3,4)	(2,3)-1=0	(8,4)-1=3)	(3,3)-1=2
(4,4)	(3,3)+2=6)		
(5,4)	(4,3)-1-2-121	(4,4)-1=4	(5,3)-1=1-1=0
(6,4)	(5,3) - 1 = 0	(5,4)-1=3	(6,3)-1=0-1=-1
(#, 4)	(6,3) + 2 = (2)	(+,4)-1=1	(8,3)-1=3-1=2
(8,4)	(4,3) -1.2	(8,4)-1=1	
(10,4)	(8,3)-1-2		(9,3)-1 = 2-1=1
	(9,3)-1-0	(9,4)-1=1	(10,3)-1=1-1=0
(1,5)	(0,4)-12-1	(0,5)-1=-1	(1,4)-1=6
(2,5)	(1,4) +2 = 3		
(3,5)	(2,4)-1=(3)	(2.5)-1= 2	(3,4)-1=2
(4,5)	(3,4)+2=65		
(5,5)	(4,4)-1=9	(4,5)-1=4	(5,4)-123
(6,5)	(3, 4) - 1 = (3)	(5,5)-1=3	(6, 4)-122
(4,5)	(6, u) + 2 = 5		
(8,t)	(7, W)-1= 1	(7,5)-1=4	(8,4)-121
(9,5)		(8,5)-1=3	(9,4)-1-1
(10,5)	(8,4)-1=1		
14,37	$(9, \omega) - 1 = 1$	(9,5)-1-(2)	(10,4)-1=0

	P, (1-1)(1-1)	D C NA	7/200-1
(1,6)	(0,5)-1:(-)	P2 (1-1)(j)	P3 (1)(3-1)
(2,6)		(0,6)-11-1	(1,5)-12-1
(3,6)	(1,5)-1:-1	(1,6)-1,-2	(2,5)-1,2
(4,6)	(2,5)-1.(2)	(2,6)-1, 1	(3,5)-1.2
(5,6)	(3,5)-1,2	(3.6)-1.1	(4,5)-1.4)
(6,6)	(6.5) - 1: (1)	(4,6)-123	
(1,6)	(5,5)+2.6		(5,5)-1=3
(8,6)	(\$,5)-1,2	(6,6)-1(5)	(4,5)-1.4
(5,6)	(\$,5)-1, (4)	(7,6)-1:4	(8,5)-1-3
(10,6)	(8,5)-1:3	(8,6)-1:3	(9,5)-1.2
(1,1)	(9,5)-1-2	(9,6)-1=2	
(2,7)	(0,6)-1=(-)	(0,7)-1=-1	(10,5)-1=1
(3,7)	$(1,6)^{-1} \cdot -2$ $(2,6)^{-1} \cdot (1)$	(1,7)-1=-2	(2,6)-12(1)
(4,7)	(3,6) -1 = 1	(2,7)-1= 0 (3,7)-1= 0	(3,6)-12 1
(5,7)	(4,6)+2-6	& -	(4,6)-1-3
(6,7)	(5,6)-1=3	(5,7)-1= (5)	
(4,7)	(6,6)-12(5)	(6,7)-1-4	(6,6)-1=5
(8,7)	(7,4)+2 = 7	(0) // /- 9	(7,6)-1=4
(9,7)	(8,6) + 2= 6		
(6,7)	(9,6) +2 · (5)	-	
(3,1)	(0,7)+2:2	1.2	
(2,8)	(1,7)-1z-2	(1,8)-1:	(2,7)-1=0
(3,8)	(2,7)+2, $(3)$	(3,8)-1-(2)	
(4,8) (5,8)	(3,7) - 12 0 (4,7) - 12 2	(9,8)-1=1	(4,7)-1:2 (5,7)-1:5
6,8)	(5,7)-1=5	(5,8)-1=4	(6,7)-1:4
4,8)	(6,7)-1=(4)	(6,8)-1=4	
6,8)	(7,7)-12 A	(4,8)-1=3	(7,4)-1-4
3,8)	(8,7)-1:6	(8,8)-1-5	(8,7)-1-6
10,8)	(9,7)-1=5	(9,8)-1.5	(9,7)-1=5
			(10,7)-1=4

(1,9) $(2,9)$ $(3,9)$ $(4,9)$ $(5,9)$ $(5,9)$ $(8,9)$ $(9,9)$ $(10,9)$	$P_{1}(9-1)(j-1)$ $(0,8)-1:-1$ $(1,8)-1:0$ $(2,8)-1:0$ $(3,8)-1:2$ $(4,8)+2:4$ $(5,8)-1:4$ $(6,8)-1:4$ $(6,8)+2:6$ $(8,8)+2:6$ $(8,8)+2:8$	$P_{2}(i-1)(j)$ $(0, 9)-1=-1$ $(1, 9)-1=0$ $(2, 9)-1=0$ $(3, 9)-1=1$ $(5, 9)-1=3$ $(6, 9)-1=3$	$P_3(1)(j-1)$ $(1,8)-1.0$ $(2,8)-1.0$ $(3,8)-1.2$ $(4,8)-1.1$ $(5,8)-1.2$ $(4,8)-1.3$
(1, 10)	-	-1	0
(2,10)		-1	0
(3,10)	0	-1	
(4,10)		0	
(5,10)		May a series of the series of	
(6,10)	A	2	<u>#</u>
(7,10)	(3)	2	3
(8,10)	(3)	2 .	
(9,10)	\$5	A	(3)
(10, 10)	7		( <del>+</del> )
			*