KDD II – Exercise 6

6.1 Longest Common Subsequences (on Time Series)

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0					
A	0					
C	0					

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0				
A	0					
C	0					

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

LCS

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**			
A	0					
C	0					

G

New match:
arrow always to left-upper diagonal cell!

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1		
A	0					
C	0					

LCS

G

No new match:
arrow to the left
and/or upper cell
(with highest value)

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

LCS

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1	1	
A	0					
C	0					

G

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

LCS

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1	1	1
A	0					
C	0					

G

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1	1	1
A	0	1				
C	0					

LCS

G

A

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1	1	1
A	0	1	1			
C	0					

LCS

G

A, G

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1	1	1
A	0	1	1	1		
C	0					

LCS

G

A, G

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1	1	1
A	0	1	1	1	**2**	
C	0					

LCS

G

GA

only GA possible, not AA

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

LCS

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1	1	1
A	0	1	1	1	**2**	2
C	0					

G

GA

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1	1	1
A	0	1	1	1	**2**	2
C	0	1				

LCS

G

GA

A

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1	1	1
A	0	1	1	1	**2**	2
C	0	1	1			

LCS

G

GA

A, G

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1	1	1
A	0	1	1	1	**2**	2
C	0	1	1	2		

LCS

G

GA

AC, GC

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1	1	1
A	0	1	1	1	**2**	2
C	0	1	1	**2**	2	

LCS

G

GA

AC, GC, GA

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1	1	1
A	0	1	1	1	**2**	2
C	0	1	1	**2**	2	2

LCS

G

GA

AC, GC, GA

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1	1	1
A	0	1	1	1	**2**	2
C	0	1	1	**2**	2	2

Traceback of LCS GA

Example LCS

X = (A, G, C, A, T) Y = (G, A, C)

	0	**A**	**G**	**C**	**A**	**T**
0	0	0	0	0	0	0
G	0	0	**1**	1	1	1
A	0	1	1	1	**2**	2
C	0	1	1	**2**	2	2

Traceback of LCS GC

LCS for Time Series

Given two time series $X = (x_1, \ldots, x_n)$ and $Y = (y_1, \ldots, y_m)$ and two threshold ϵ and δ,

$$LSC(X,Y) = \begin{cases} 0 & if\ n = 0 \vee m = 0 \\ LCS(start(X), start(Y)) + 1 & if\ match(last(X), last(Y)) \\ max(LCS(start(X), Y), LCS(X, start(Y))) & else \end{cases}$$

where the matching function is defined as:

$$match(x_i, y_j) = \begin{cases} true & if\ |x_i - y_j| \leq \epsilon \wedge |i - j| \leq \delta \\ false & else \end{cases}$$

Then the distance is:

$$D_{LCS}(X,Y) = 1 - \frac{LCS(X,Y)}{\min(n,m)}$$

		1	2	3	4	5	6	7
	0	**3**	**5**	**9**	**2**	**3**	**6**	**3**
0	0	0	0	0	0	0	0	0
1 **3**	0							
2 **4**	0							
3 **6**	0							
4 **10**	0							
5 **1**	0							
6 **3**	0							
7 **2**	0							
8 **7**	0							
9 **4**	0							
10								

$\delta = 2$

$\epsilon = 2$

		1	2	3	4	5	6	7
	0	**3**	**5**	**9**	**2**	**3**	**6**	**3**
0	0	0	0	0	0	0	0	0
3	0	1						
4	0							
6	0							
10	0							
1	0							
3	0							
2	0							
7	0							
4	0							

1
2
3
4
5
6
7
8
9
10

$\delta = 2$

$\epsilon = 2$

		1	2	3	4	5	6	7
	0	**3**	**5**	**9**	**2**	**3**	**6**	**3**
0	0	0	0	0	0	0	0	0
1 **3**	0	1	1					
2 **4**	0							
3 **6**	0							
4 **10**	0							
5 **1**	0							
6 **3**	0							
7 **2**	0							
8 **7**	0							
9 **4**	0							
10								

$\delta = 2$

$\epsilon = 2$

		1	2	3	4	5	6	7
	0	**3**	**5**	**9**	**2**	**3**	**6**	**3**
0	0	0	0	0	0	0	0	0
1 **3**	0	1	1	1				
2 **4**	0							
3 **6**	0							
4 **10**	0							
5 **1**	0							
6 **3**	0							
7 **2**	0							
8 **7**	0							
9 **4**	0							
10								

$\delta = 2$

$\epsilon = 2$

		1	2	3	4	5	6	7
	0	**3**	**5**	**9**	**2**	**3**	**6**	**3**
0	0	0	0	0	0	0	0	0
1 **3**	0	1	1	1				
2 **4**	0	1						
3 **6**	0							
4 **10**	0							
5 **1**	0							
6 **3**	0							
7 **2**	0							
8 **7**	0							
9 **4**	0							
10								

$\delta = 2$

$\epsilon = 2$

		1	2	3	4	5	6	7
	0	**3**	**5**	**9**	**2**	**3**	**6**	**3**
0	0	0	0	0	0	0	0	0
3	0	1	1	1				
4	0	1	2					
6	0							
10	0							
1	0							
3	0							
2	0							
7	0							
4	0							

1
2
3
4
5
6
7
8
9
10

$\delta = 2$

$\epsilon = 2$

		1	2	3	4	5	6	7
	0	**3**	**5**	**9**	**2**	**3**	**6**	**3**
0	0	0	0	0	0	0	0	0
1 **3**	0	1	1	1				
2 **4**	0	1	2	2				
3 **6**	0							
4 **10**	0							
5 **1**	0							
6 **3**	0							
7 **2**	0							
8 **7**	0							
9 **4**	0							
10								

$\delta = 2$

$\epsilon = 2$

		1	2	3	4	5	6	7
	0	**3**	**5**	**9**	**2**	**3**	**6**	**3**
0	0	0	0	0	0	0	0	0
3	0	1	1	1				
4	0	1	2	2	2			
6	0							
10	0							
1	0							
3	0							
2	0							
7	0							
4	0							

1
2
3
4
5
6
7
8
9
10

$\delta = 2$

$\epsilon = 2$

		1	2	3	4	5	6	7
	0	**3**	**5**	**9**	**2**	**3**	**6**	**3**
0	0	0	0	0	0	0	0	0
3	0	1	1	1				
4	0	1	2	2	2			
6	0	1						
10	0							
1	0							
3	0							
2	0							
7	0							
4	0							

1
2
3
4
5
6
7
8
9
10

$\delta = 2$

$\epsilon = 2$

		1	2	3	4	5	6	7
	0	**3**	**5**	**9**	**2**	**3**	**6**	**3**
0	0	0	0	0	0	0	0	0
3	0	1	1	1				
4	0	1	2	2	2			
6	0	1	2					
10	0							
1	0							
3	0							
2	0							
7	0							
4	0							

1
2
3
4
5
6
7
8
9
10

$\delta = 2$

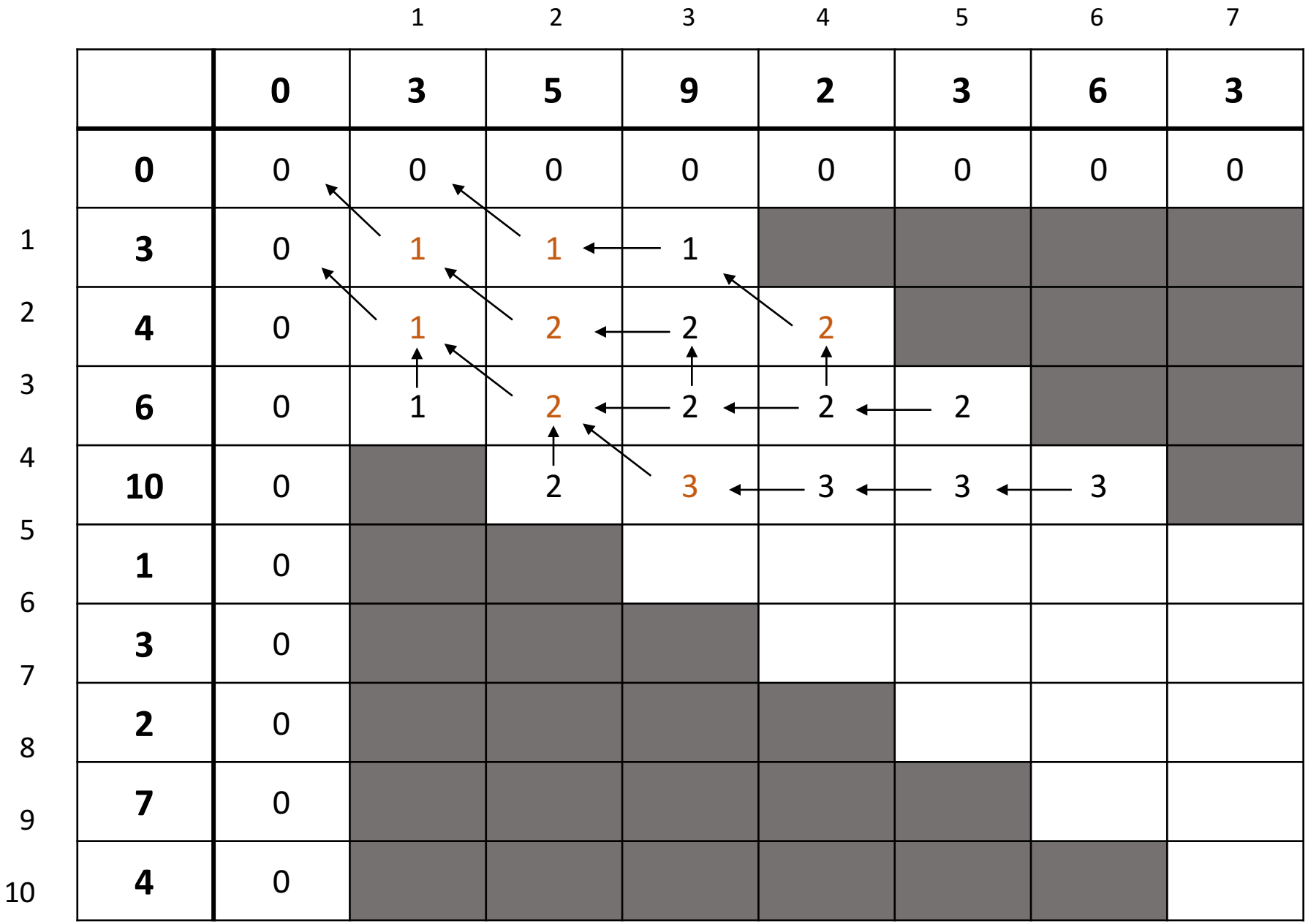
$\epsilon = 2$

		1	2	3	4	5	6	7
	0	**3**	**5**	**9**	**2**	**3**	**6**	**3**
0	0	0	0	0	0	0	0	0
3	0	1	1	1				
4	0	1	2	2	2			
6	0	1	2	2	2	2		
10	0							
1	0							
3	0							
2	0							
7	0							
4	0							

1
2
3
4
5
6
7
8
9
10

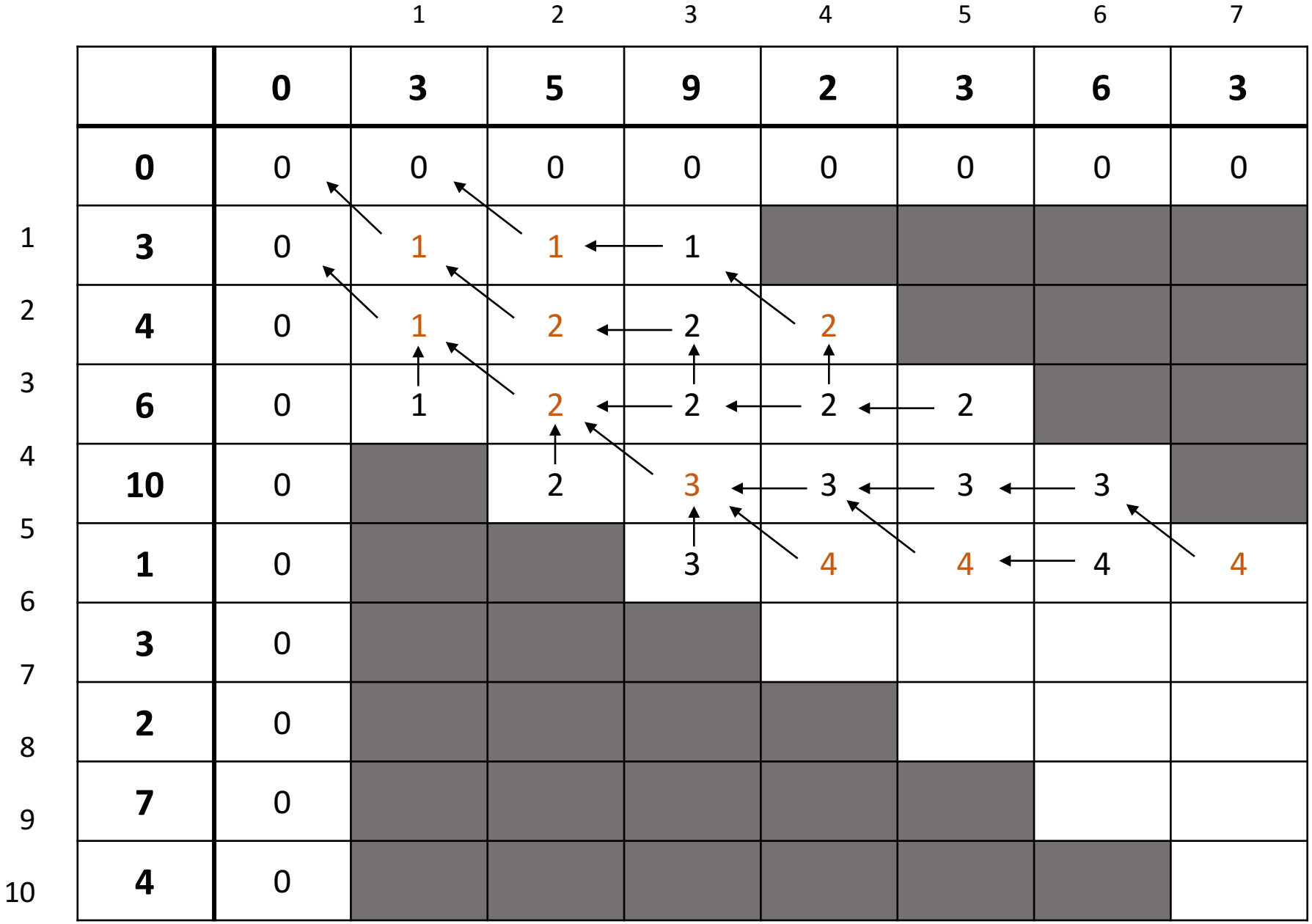
$\delta = 2$

$\epsilon = 2$



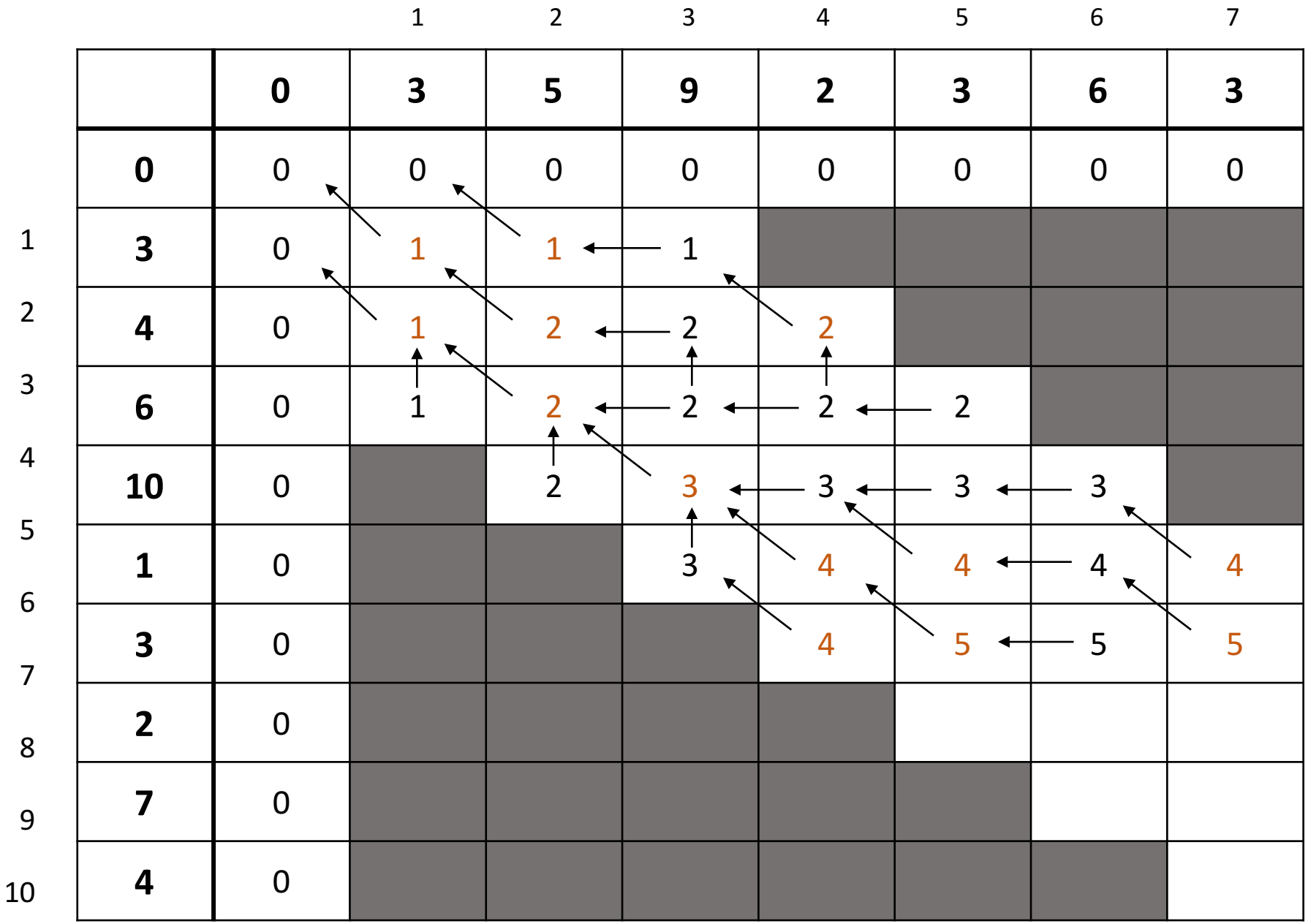
$\delta = 2$

$\epsilon = 2$



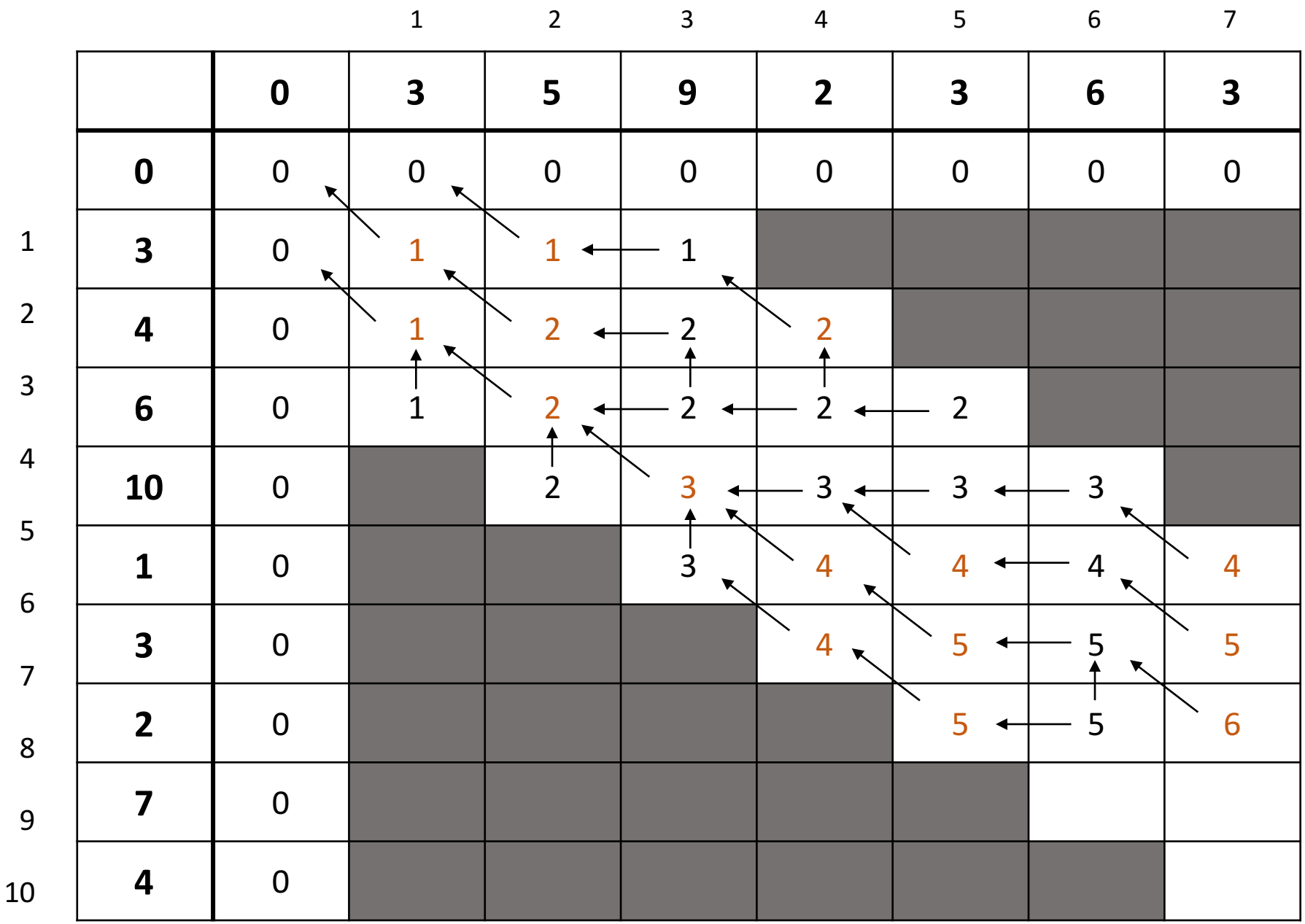
$\delta = 2$

$\epsilon = 2$



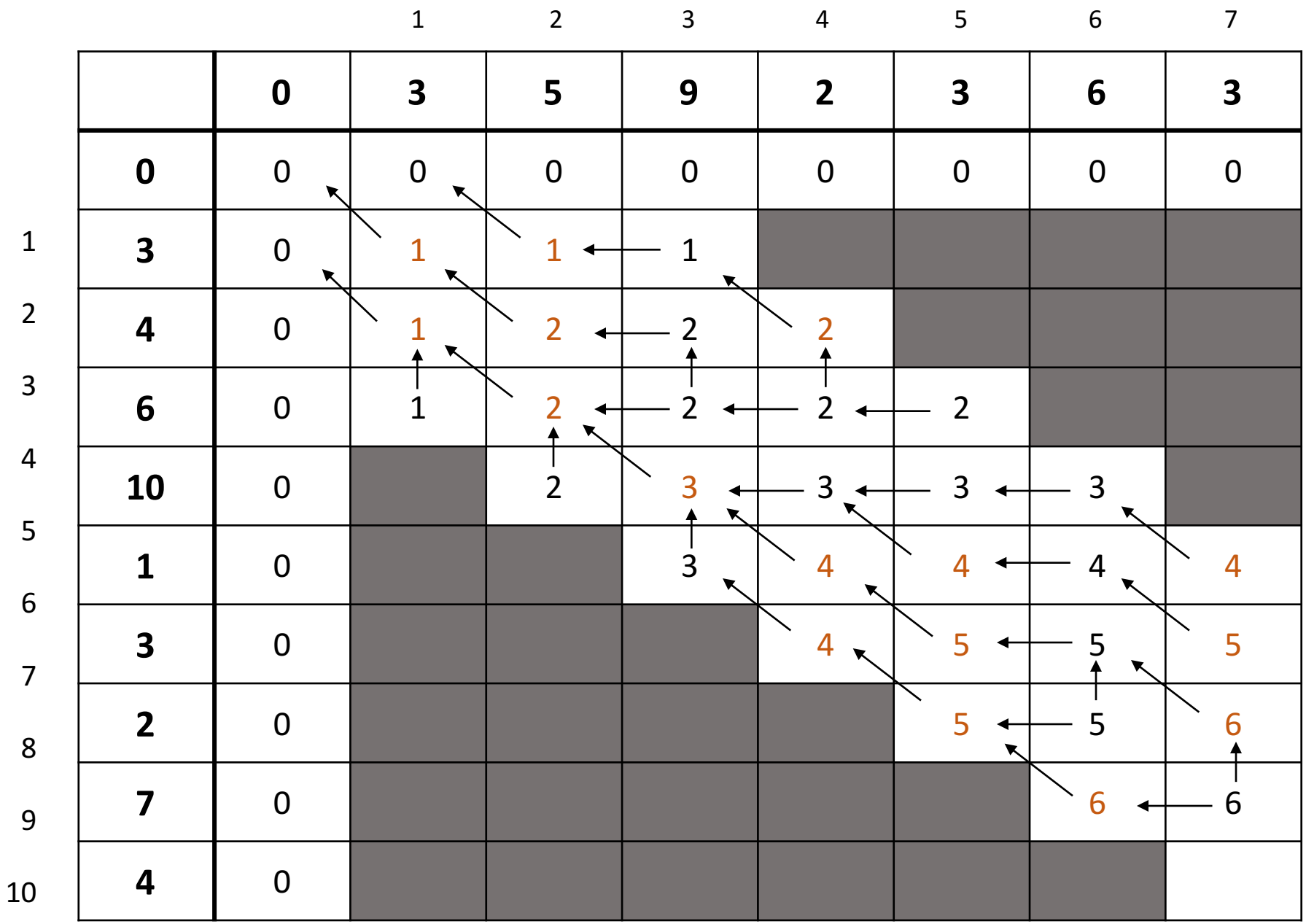
$\delta = 2$

$\epsilon = 2$



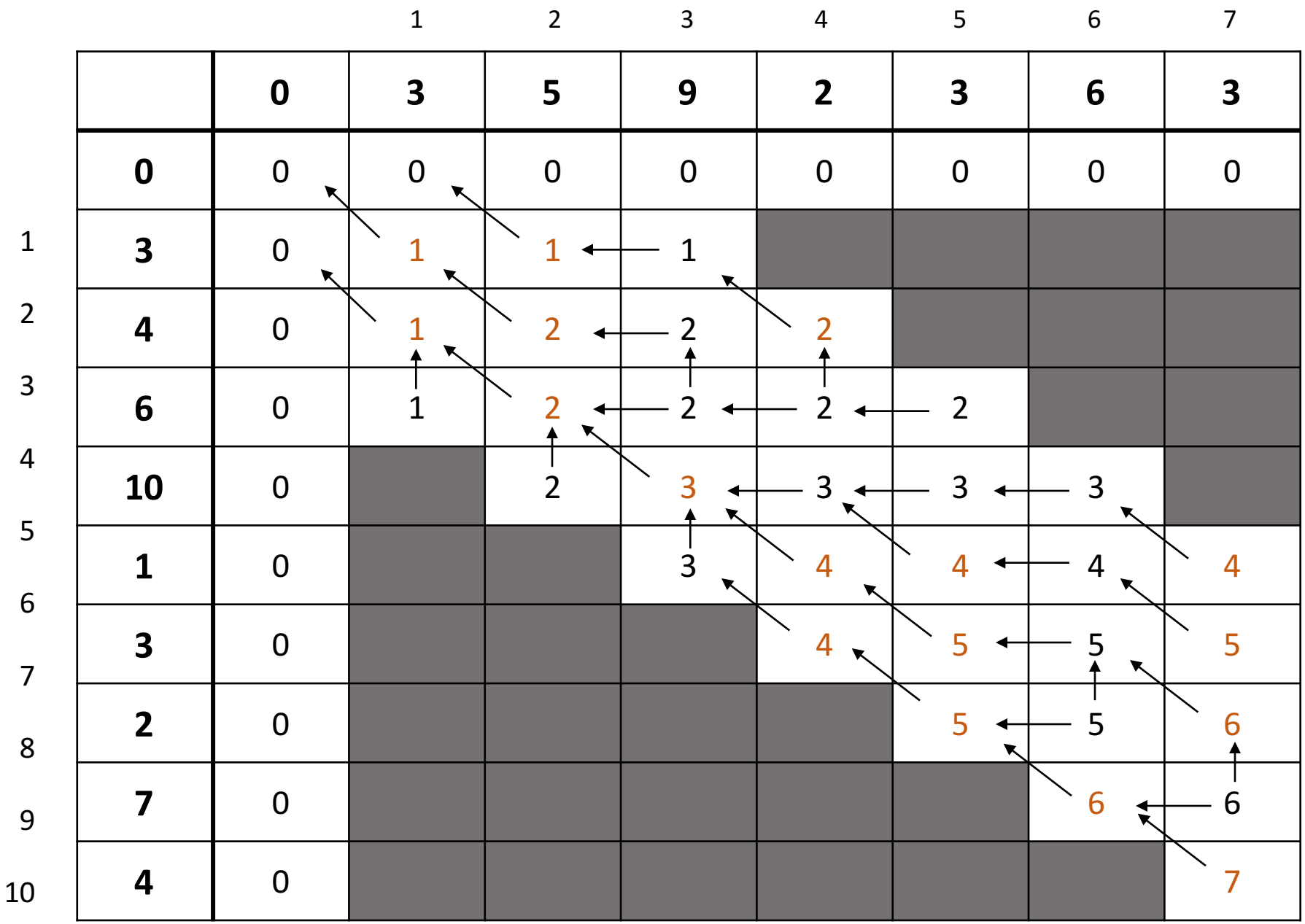
$\delta = 2$

$\epsilon = 2$



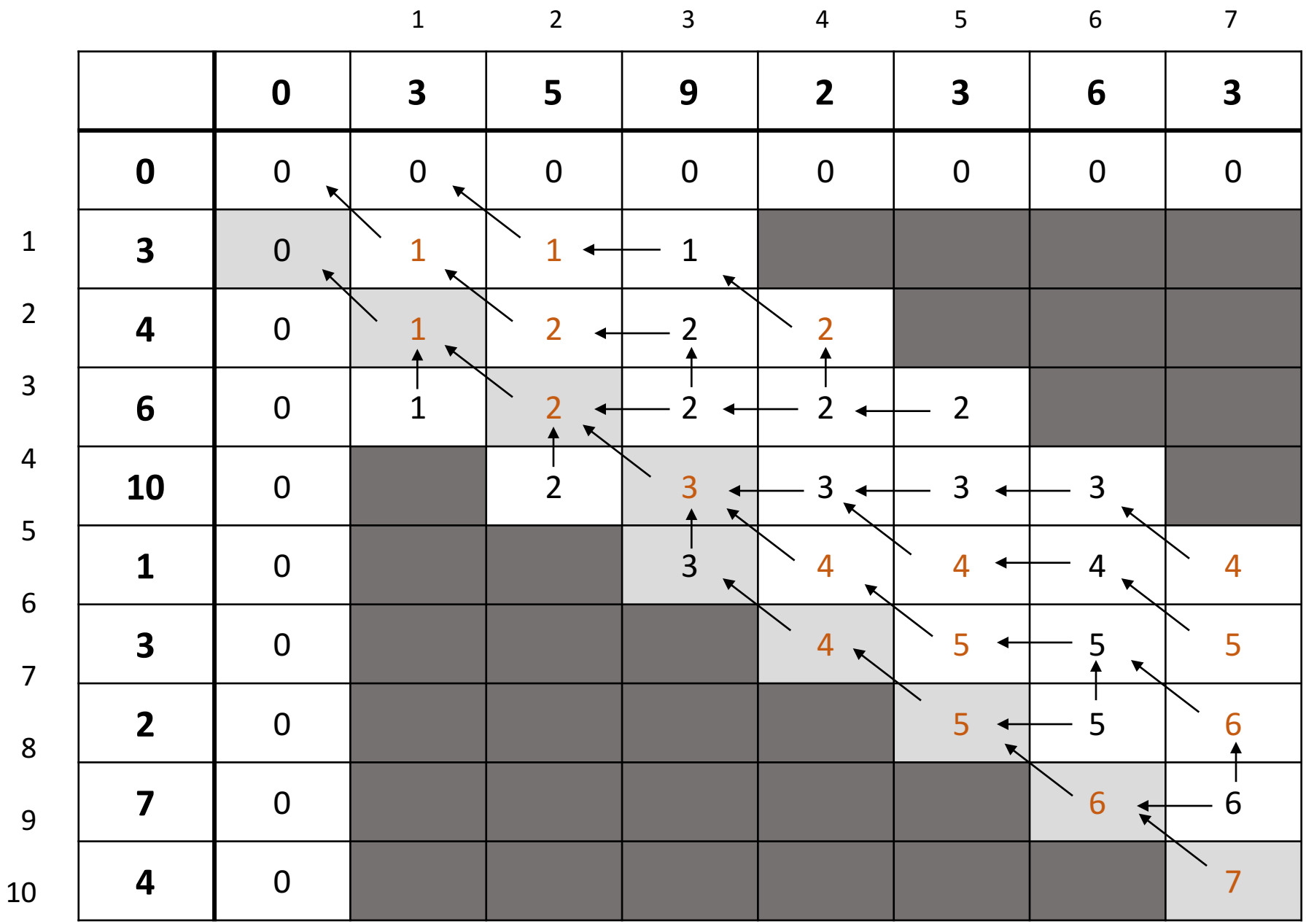
$\delta = 2$

$\epsilon = 2$



$\delta = 2$

$\epsilon = 2$



$\delta = 2$

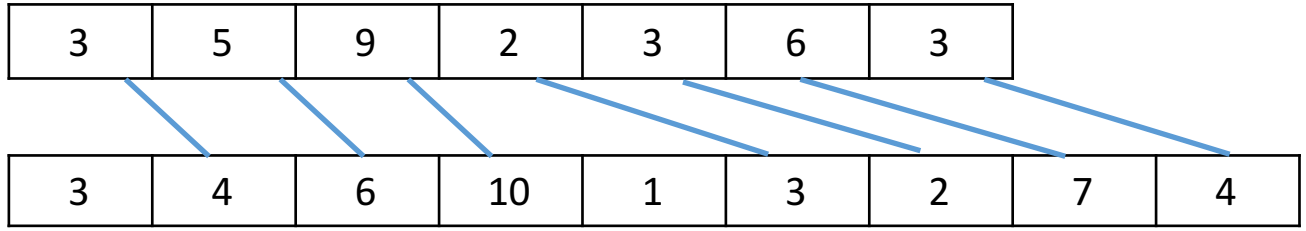
$\epsilon = 2$

	0	**3**	**5**	**9**	**2**	**3**	**6**	**3**
0	0	0	0	0	0	0	0	0
3	0	1	1	1				
4	0	1	2	2	2			
6	0	1	2	2	2	2		
10	0		2	3	3	3	3	
1	0			3	4	4	4	4
3	0				4	5	5	5
2	0					5	5	6
7	0						6	6
4	0							7

$$\delta = 2$$

$$\epsilon = 2$$

Matching



LCS (X, Y) = 7