# Key steps to design DP algorithms

1. Identify subproblems

# Key steps to design DP algorithms

1. Identify subproblems
2. Recurrence

   e.g. $L(j) = 1 + \max\{L(i) : a_i < a_j\}$

# Key steps to design DP algorithms

1. Identify subproblems
2. Recurrence

   e.g. $L(j) = 1 + \max\{L(i) : a_i < a_j\}$
3. Base case

# Dynamic Programming

## Edit Distance (Textbook Section 6.3)

# Edit distance

Motivation: consider DNA sequences $x = ACGTA$, $y = ATCTG$.

# Edit distance

Motivation: consider DNA sequences $x = ACGTA$, $y = ATCTG$.

Note $|x| \neq |y|$ in general

# Edit distance

Motivation: consider DNA sequences $x = ACGTA$, $y = ATCTG$.

Note $|x| \neq |y|$ in general

Question: how far away are $x$ and $y$?

# Edit distance

Motivation: consider DNA sequences $x = ACGTA$, $y = ATCTG$.

Note $|x| \neq |y|$ in general

Question: how far away are $x$ and $y$?

**Definition**

The **edit distance** between $x$ and $y$, denoted by $d(x, y)$, is the minimum number of insertions, deletions, and substitutions needed to transform $x$ to $y$

## Edit distance

Motivation: consider DNA sequences $x = ACGTA$, $y = ATCTG$.

Note $|x| \neq |y|$ in general

Question: how far away are $x$ and $y$?

**Definition**

The **edit distance** between $x$ and $y$, denoted by $d(x, y)$, is the minimum number of insertions, deletions, and substitutions needed to transform $x$ to $y$

Consider the following **alignments**:

# Edit distance

Motivation: consider DNA sequences $x = ACGTA$, $y = ATCTG$.

Note $|x| \neq |y|$ in general

Question: how far away are $x$ and $y$?

**Definition**

The **edit distance** between $x$ and $y$, denoted by $d(x, y)$, is the minimum number of insertions, deletions, and substitutions needed to transform $x$ to $y$

Consider the following **alignments**:

| | | | | | | |
|---|---|---|---|---|---|---|
| x: | A | - | C | G | T | A |
| y: | A | T | C | - | T | G |

Cost : 3

# Edit distance

Motivation: consider DNA sequences $x = ACGTA$, $y = ATCTG$.

Note $|x| \neq |y|$ in general

Question: how far away are $x$ and $y$?

**Definition**

The **edit distance** between $x$ and $y$, denoted by $d(x, y)$, is the minimum number of insertions, deletions, and substitutions needed to transform $x$ to $y$

Consider the following **alignments**:

| | | | | | | |
|---|---|---|---|---|---|---|
| x: | A | - | C | G | T | A |
| y: | A | T | C | - | T | G |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| x: | A | C | - | - | G | T | A |
| y: | A | T | C | T | G | - | - |

cost: 5

# Edit distance

Motivation: consider DNA sequences $x = ACGTA$, $y = ATCTG$.

Note $|x| \neq |y|$ in general

Question: how far away are $x$ and $y$?

**Definition**

The **edit distance** between $x$ and $y$, denoted by $d(x, y)$, is the minimum number of insertions, deletions, and substitutions needed to transform $x$ to $y$

Consider the following **alignments**:

| | | | | | | |
|---|---|---|---|---|---|---|
| x: | A | - | C | G | T | A |
| y: | A | T | C | - | T | G |

$\text{cost} : 3$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| x: | A | C | - | - | G | T | A |
| y: | A | T | C | T | G | - | - |

## Edit distance

Motivation: consider DNA sequences $x = ACGTA$, $y = ATCTG$.

Note $|x| \neq |y|$ in general

Question: how far away are $x$ and $y$?

**Definition**

The **edit distance** between $x$ and $y$, denoted by $d(x, y)$, is the minimum number of insertions, deletions, and substitutions needed to transform $x$ to $y$

Consider the following **alignments**:

| | | | | | | |
|---|---|---|---|---|---|---|
| x: | A | - | C | G | T | A |
| y: | A | T | C | - | T | G |

cost : 3

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| x: | A | C | - | - | G | T | A |
| y: | A | T | C | T | G | - | - |

cost : 5

# Edit distance

Motivation: consider DNA sequences $x = ACGTA$, $y = ATCTG$.

Note $|x| \neq |y|$ in general

Question: how far away are $x$ and $y$?

**Definition**

The **edit distance** between $x$ and $y$, denoted by $d(x, y)$, is the minimum number of insertions, deletions, and substitutions needed to transform $x$ to $y$

Consider the following **alignments**:

| | | | | | | |
|---|---|---|---|---|---|---|
| x: | A | - | C | G | T | A |
| y: | A | T | C | - | T | G |

cost : 3 (optimal)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| x: | A | C | - | - | G | T | A |
| y: | A | T | C | T | G | - | - |

cost : 5

# Edit distance

Motivation: consider DNA sequences $x = ACGTA$, $y = ATCTG$.

Note $|x| \neq |y|$ in general

Question: how far away are $x$ and $y$?

**Definition**

The **edit distance** between $x$ and $y$, denoted by $d(x, y)$, is the minimum number of insertions, deletions, and substitutions needed to transform $x$ to $y$

Consider the following **alignments**:

| | | | | | | |
|---|---|---|---|---|---|---|
| x: | A | - | C | G | T | A |
| y: | A | T | C | - | T | G |

cost : 3 (optimal)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| x: | A | C | - | - | G | T | A |
| y: | A | T | C | T | G | - | - |

cost : 5

So $d(x, y) = 3$

# Edit distance — subproblem

Consider two strings

length: m length: n

$$x = x_1 x_2 \cdots x_m \quad \text{and} \quad y = y_1 y_2 \cdots y_n$$

# Edit distance — subproblem

Consider two strings

$$x = x_1 x_2 \cdots x_m \quad \text{and} \quad y = y_1 y_2 \cdots y_n$$

Subproblem: consider prefix $x_1 \cdots x_i$ and $y_1 \cdots y_j$ ($i \leq m$, $j \leq n$)

# Edit distance — subproblem

$d(x, y)$

Consider two strings

$$x = x_1 x_2 \cdots x_m \quad \text{and} \quad y = y_1 y_2 \cdots y_n$$

Subproblem: consider prefix $x_1 \cdots x_i$ and $y_1 \cdots y_j$ ($i \le m$, $j \le n$)

Define

$$E(i, j) = d(x_1 \cdots x_i, y_1 \cdots y_j)$$

# Edit distance — subproblem

Consider two strings

$$x = x_1 x_2 \cdots x_m \quad \text{and} \quad y = y_1 y_2 \cdots y_n$$

Subproblem: consider prefix $x_1 \cdots x_i$ and $y_1 \cdots y_j$ ($i \leq m$, $j \leq n$)

Define

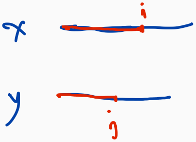
$$E(i,j) = d(x_1 \cdots x_i, y_1 \cdots y_j)$$

Optimal solution: $E(m,n)$

$$\boxed{E(m,n)} = d(\underbrace{x_1 \cdots x_m}_{x}, \underbrace{y_1 \cdots y_n}_{y})$$

$$= d(x,y)$$

# Edit distance — subproblem




Consider two strings

$$x = x_1 x_2 \cdots x_m \quad \text{and} \quad y = y_1 y_2 \cdots y_n$$

Subproblem: consider prefix $x_1 \cdots x_i$ and $y_1 \cdots y_j$ ($i \le m$, $j \le n$)

Define

$$E(i,j) = d(x_1 \cdots x_i, y_1 \cdots y_j)$$

$E(i, 0)$

$= d(x_1 \cdots x_i, -) = i$

Optimal solution: $E(m,n)$

How to use the solution to the subproblems to solve $E(i,j)$?

# Recurrence (I)

Look at the rightmost column:

Look at the rightmost column:

Case 1 $\begin{bmatrix} x_1 & \cdots & x_{i-1} \\ y_1 & \cdots & y_j \end{bmatrix} \begin{bmatrix} \color{red}{x_i} \\ \color{red}{-} \end{bmatrix}$

$E(i,j) = 1 + E(i-1, j)$

# Recurrence (I)

Look at the rightmost column:

**Case 1** $\begin{array}{cccc} x_1 & \cdots & x_{i-1} & x_i \\ y_1 & \cdots & y_j & - \end{array}$

Contributes 1 to the cost plus the cost of alignment $\begin{array}{ccc} x_1 & \cdots & x_{i-1} \\ y_1 & \cdots & y_j \end{array}$

# Recurrence (I)

Look at the rightmost column:

**Case 1** $\begin{array}{cccc} x_1 & \cdots & x_{i-1} & x_i \\ y_1 & \cdots & y_j & - \end{array}$

Contributes 1 to the cost plus the cost of alignment $\begin{array}{ccc} x_1 & \cdots & x_{i-1} \\ y_1 & \cdots & y_j \end{array}$

$$E(i,j) = 1 + \boxed{E(i-1,j)}$$

$\begin{array}{ccccc} x_1 & --- & x_{i-2} & x_{i-1} & x_i \\ y_1 & ---- & y_j & - & - \end{array}$

$E(i-1, j)$

## Recurrence (I)

Look at the rightmost column:

Case 1 $\begin{array}{cccc} x_1 & \cdots & x_{i-1} & x_i \\ y_1 & \cdots & y_j & - \end{array}$

Contributes 1 to the cost plus the cost of alignment $\begin{array}{ccc} x_1 & \cdots & x_{i-1} \\ y_1 & \cdots & y_j \end{array}$

$$E(i,j) = 1 + E(i-1,j)$$

Case 2 $\begin{array}{cccc} x_1 & \cdots & x_i & - \\ y_1 & \cdots & y_{j-1} & y_j \end{array}$

Look at the rightmost column:

Case 1 $\begin{array}{cccc} x_1 & \cdots & x_{i-1} & x_i \\ y_1 & \cdots & y_j & - \end{array}$

Contributes 1 to the cost plus the cost of alignment $\begin{array}{ccc} x_1 & \cdots & x_{i-1} \\ y_1 & \cdots & y_j \end{array}$

$$E(i,j) = 1 + E(i-1,j)$$

Case 2 $\begin{array}{cccc} x_1 & \cdots & x_i & - \\ y_1 & \cdots & y_{j-1} & y_j \end{array}$

Contributes 1 to the cost plus the cost of alignment $\begin{array}{ccc} x_1 & \cdots & x_i \\ y_1 & \cdots & y_{j-1} \end{array}$

## Recurrence (I)

Look at the rightmost column:

**Case 1**

$$\begin{matrix} x_1 & \cdots & x_{i-1} & x_i \\ y_1 & \cdots & y_j & \text{-} \end{matrix}$$

Contributes 1 to the cost plus the cost of alignment $\begin{matrix} x_1 & \cdots & x_{i-1} \\ y_1 & \cdots & y_j \end{matrix}$

$$E(i,j) = 1 + E(i-1,j)$$

**Case 2**

$$\begin{matrix} x_1 & \cdots & x_i & \text{-} \\ y_1 & \cdots & y_{j-1} & y_j \end{matrix}$$

Contributes 1 to the cost plus the cost of alignment $\begin{matrix} x_1 & \cdots & x_i \\ y_1 & \cdots & y_{j-1} \end{matrix}$

$$E(i,j) = 1 + E(i,j-1)$$

# Recurrence (I)

Look at the rightmost column:

**Case 1** $\begin{matrix} x_1 & \cdots & x_{i-1} & \color{red}{x_i} \\ y_1 & \cdots & y_j & \color{red}{-} \end{matrix}$

Contributes 1 to the cost plus the cost of alignment $\begin{matrix} x_1 & \cdots & x_{i-1} \\ y_1 & \cdots & y_j \end{matrix}$

$$E(i, j) = 1 + E(i-1, j)$$

**Case 2** $\begin{matrix} x_1 & \cdots & x_i & \color{red}{-} \\ y_1 & \cdots & y_{j-1} & \color{red}{y_j} \end{matrix}$

Contributes 1 to the cost plus the cost of alignment $\begin{matrix} x_1 & \cdots & x_i \\ y_1 & \cdots & y_{j-1} \end{matrix}$

$$E(i, j) = 1 + E(i, j-1)$$

**Case 3** $\begin{matrix} x_1 & \cdots & x_{i-1} & \color{red}{x_i} \\ y_1 & \cdots & y_{j-1} & \color{red}{y_j} \end{matrix}$

# Recurrence (I)

Look at the rightmost column:

**Case 1**
$$\begin{array}{cccc} x_1 & \cdots & x_{i-1} & \color{red}{x_i} \\ y_1 & \cdots & y_j & \color{red}{-} \end{array}$$

Contributes 1 to the cost plus the cost of alignment $\begin{array}{ccc} x_1 & \cdots & x_{i-1} \\ y_1 & \cdots & y_j \end{array}$

$$E(i,j) = 1 + E(i-1,j)$$

**Case 2**
$$\begin{array}{cccc} x_1 & \cdots & x_i & \color{red}{-} \\ y_1 & \cdots & y_{j-1} & \color{red}{y_j} \end{array}$$

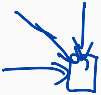
Contributes 1 to the cost plus the cost of alignment $\begin{array}{ccc} x_1 & \cdots & x_i \\ y_1 & \cdots & y_{j-1} \end{array}$

$E(i,j)$

$$E(i,j) = 1 + E(i,j-1)$$

**Case 3**
$$\begin{array}{cccc} x_1 & \cdots & x_{i-1} & \color{red}{x_i} \\ y_1 & \cdots & y_{j-1} & \color{red}{y_j} \end{array}$$

$$E(i,j) = \begin{cases} E(i-1,j-1) & \text{if } x_i = y_j \\ 1 + E(i-1,j-1) & \text{otherwise} \end{cases}$$



**The recurrence:**

$$\boxed{E(i,j) = \min\{1 + E(i-1,j), 1 + E(i,j-1), \mathrm{diff}(i,j) + E(i-1,j-1)\},}$$

where

$$\mathrm{diff}(i,j) = \begin{cases} 1 & \text{if } x_i \neq y_j \\ 0 & \text{otherwise} \end{cases}$$

E 0 1 2 3 4 ......

0

1 E(1,2)

2

3

4

⋮

# Recurrence (II)

**The recurrence:**

$$E(i,j) = \min\{1 + E(i-1,j), 1 + E(i,j-1), \mathrm{diff}(i,j) + E(i-1,j-1)\},$$

where

$$\mathrm{diff}(i,j) = \begin{cases} 1 & \text{if } x_i \neq y_j \\ 0 & \text{otherwise} \end{cases}$$

**Optimal solution:** $E(m,n)$

# Recurrence (II)

**The recurrence:**

$$E(i,j) = \min\{1 + E(i-1,j), 1 + E(i,j-1), \mathrm{diff}(i,j) + E(i-1,j-1)\},$$

where

$$\mathrm{diff}(i,j) = \begin{cases} 1 & \text{if } x_i \neq y_j \\ 0 & \text{otherwise} \end{cases}$$

**Optimal solution:** $E(m,n)$

**Base case:** $E(0,0) = 0$, $E(i,0) = i$, $E(0,j) = j$

# Filling the table

$$E(i,j) = \min\{1 + E(i-1,j), 1 + E(i,j-1), \mathrm{diff}(i,j) + E(i-1,j-1)\},$$

# Filling the table

$$E(i,j) = \min\{1 + E(i-1,j), 1 + E(i,j-1), \mathrm{diff}(i,j) + E(i-1,j-1)\},$$

$$\begin{array}{cccccccc}
E(0,0) & & E(0,1) & \cdots & & E(0,n-1) & & E(0,n) \\
 & \searrow & \downarrow & & & & & \\
E(1,0) & \rightarrow & E(1,1) & \cdots & & \cdots & & \cdots \\
\vdots & & & & & & & \\
 & & & & \searrow & \downarrow & \searrow & \downarrow \\
E(m-1,0) & & & & \rightarrow & E(m-1,n-1) & \rightarrow & E(m-1,n) \\
 & & & & \searrow & \downarrow & \searrow & \downarrow \\
E(m,0) & & & & \rightarrow & E(m,n-1) & \rightarrow & E(m,n)
\end{array}$$

## Running example

$x = ACGTA$ and $y = ATCTG$

# Running example

$x = ACGTA$ and $y = ATCTG$

| | | $A$ | $T$ | $C$ | $T$ | $G$ |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 +1 = 3 | 3 +1 = 4 | 4 | 5 |
| A | 1 | 0 | 1 +1 = 2 | 2 | | |
| C | 2 | | | | | |
| G | 3 | | | | | |
| T | 4 | | | | | |
| A | 5 | | | | | |

D