Other Dynamic Programming Problems

View in slide-show mode
Problem 1
Subset Sum
Subset-Sum Problem

Given:
- a set of integers $X = \{x_1, x_2, \ldots, x_n\}$, and
- an integer $B$

Find:
- a subset of $X$ that has maximum sum not exceeding $B$.

Notation: $S_{n,B} = \{x_1, x_2, \ldots, x_n : B\}$ is the subset-sum problem
- The integers to choose from: $x_1, x_2, \ldots, x_n$
- Desired sum: $B$
Subset-Sum Problem

**Example:**

\[
x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7 \ x_8 \ x_9 \ x_{10} \ x_{11} \ x_{12}
\]

\[S_{12,99}: \{20, 30, 14, 70, 40, 50, 15, 25, 80, 60, 10, 95: 99\}\]

Find a subset of \( X \) with maximum sum not exceeding 99.

An optimal solution:

\[
x_1 \ x_3 \ x_5 \ x_8
\]

\[N_{opt} = \{20, 14, 40, 25\}\]

*with sum* \(20 + 14 + 40 + 25 = 99\)*
Optimal Substructure Property

- Consider the solution as a sequence of \( n \) decisions:
  
  \( \text{\textit{i}th \ decision} \): whether we pick number \( x_i \) or not

Let \( N_{\text{opt}} \) be an optimal solution for \( S_{n,B} \)
Let \( x_k \) be the highest-indexed number in \( N_{\text{opt}} \)

\[
N'_{\text{opt}} = N_{\text{opt}} - \{x_k\}
\]
Optimal Substructure Property

**Lemma**: $N'_{\text{opt}} = N_{\text{opt}} - \{x_k\}$ is an optimal solution for the subproblem $S_{k-1, \{B-x_k\}} = \{x_1, x_2, ..., x_{k-1}: B-x_k\}$ and

$$c(N_{\text{opt}}) = x_k + c(N'_{\text{opt}})$$

where $c(N)$ is the sum of all numbers in subset $N$
Optimal Substructure Property - Proof

Proof: By contradiction, assume that there exists another solution $A'$ for $S_{k-1, B-x_k}$ for which:

$$c(A') > c(N'_\text{opt}) \text{ and } c(A') \leq B - x_k$$

i.e. $A'$ is a better solution than $N'_\text{opt}$ for $S_{k-1, B-xk}$

Then, we can construct $A = A' \cup \{x_k\}$ as a solution to $S_{k, B}$. We have:

$$c(A) = c(A') + x_k$$

$$> c(N'_\text{opt}) + x_k = c(N_{\text{opt}})$$

Contradiction! $N_{\text{opt}}$ was assumed to be optimal for $S_{k,B}$.

Proof complete.
Optimal Substructure Property - Example

**Example:**

\[
x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7 \ x_8 \ x_9 \ x_{10} \ x_{11} \ x_{12}
\]

\[S_{12,99} = \{20, 30, 14, 70, 40, 50, 15, 25, 80, 60, 10, 95: 99\}\]

\[N_{\text{opt}} = \{20, 14, 40, 25\}\] is optimal for \(S_{12,99}\)

\[N'_{\text{opt}} = N_{\text{opt}} - \{x_8\} = \{20, 14, 40\}\] is optimal for

the subproblem \(S_{7,74} = \{20, 30, 14, 70, 40, 50, 15: 74\}\)

and

\[c(N_{\text{opt}}) = 25 + c(N'_{\text{opt}}) = 25 + 74 = 99\]
Recursive Definition an Optimal Solution

c[i, b]: the value of an optimal solution for \( S_{i,b} = \{x_1, \ldots, x_i : b\} \)

\[
c[i, b] = \begin{cases} 
0 & \text{if } i = 0 \text{ or } b = 0 \\
c[i - 1, b] & \text{if } x_i > b \\
\text{Max}\{x_i + c[i - 1, b - x_i], c[i - 1, b]\} & \text{if } i > 0 \text{ and } b \geq x_i 
\end{cases}
\]

According to this recurrence, an optimal solution \( N_{i,b} \) for \( S_{i,b} \):

- either contains \( x_i \) \( \Rightarrow \) \( c(N_{i,b}) = x_i + c(N_{i-1, b-x_i}) \)
- or does not contain \( x_i \) \( \Rightarrow \) \( c(N_{i,b}) = c(N_{i-1, b}) \)
\[ c[i, b] = \begin{cases} 
0 & \text{if } i = 0 \text{ or } b = 0 \\
c[i - 1, b] & \text{if } x_i > b \\
\max\{x_i + c[i - 1, b - x_i], c[i - 1, b]\} & \text{if } i > 0 \text{ and } b \geq x_i 
\end{cases} \]

Need to process: \( c[i, b] \)

after computing:
- \( c[i - 1, b] \)
- \( c[i - 1, b - x_i] \)
\[
c[i, b] = \begin{cases} 
0 & \text{if } i = 0 \text{ or } b = 0 \\
c[i - 1, b] & \text{if } x_i > b \\
\max\{x_i + c[i - 1, b - x_i], c[i - 1, b]\} & \text{if } i > 0 \text{ and } b \geq x_i 
\end{cases}
\]
Computing the Optimal Subset-Sum Value

**SUBSET-SUM** \((x, n, B)\)

\[
\begin{align*}
&\text{for } b \leftarrow 0 \text{ to } B \text{ do} \\
&\quad c[0, b] \leftarrow 0 \\
&\text{for } i \leftarrow 1 \text{ to } n \text{ do} \\
&\quad c[i, 0] \leftarrow 0 \\
&\text{for } i \leftarrow 1 \text{ to } n \text{ do} \\
&\quad \text{for } b \leftarrow 1 \text{ to } B \text{ do} \\
&\quad\quad \text{if } x_i \leq b \text{ then} \\
&\quad\quad\quad c[i, b] \leftarrow \text{Max} \{x_i + c[i-1, b-x_i], c[i-1, b]\} \\
&\quad\quad\text{else} \\
&\quad\quad\quad c[i, b] \leftarrow c[i-1, b] \\
&\text{return } c[n, B]
\end{align*}
\]
Finding an Optimal Subset

**SOLUTION-SUBSET-SUM** \((x, b, B, c)\)

\[
i \leftarrow n \hfill \\
b \leftarrow B \hfill \\
N \leftarrow \emptyset \hfill \\
\text{while } i > 0 \text{ do} \hfill \\
\quad \text{if } c[i, b] = c[i-1, b] \text{ then} \hfill \\
\quad \quad i \leftarrow i - 1 \hfill \\
\quad \text{else} \hfill \\
\quad \quad N \leftarrow N \cup \{x_i\} \hfill \\
\quad \quad i \leftarrow i - 1 \hfill \\
\quad \quad b \leftarrow b - x_i \hfill \\
\text{return } N \hfill \\
\]
Problem 2
Optimal Binary Search Tree
Reminder: Binary Search Tree (BST)

All keys in the left subtree less than 8

All keys in the right subtree greater than 8

This property holds for all nodes.

Image from Wikimedia
**Example**: English-to-French translation

Organize (English, French) word pairs in a BST

- **Keyword**: English word
- **Satellite data**: French word

We can search for an English word (node key) efficiently, and return the corresponding French word (satellite data).
Binary Search Tree Example

Suppose we know the frequency of each keyword in texts:

```
begin  do  else  end  if  then  while
5%    40%  8%    4%    10%  10%   23%
```

![Binary Search Tree Diagram]
Cost of a Binary Search Tree

Example: If we search for keyword “while”, we need to access 3 nodes. So, 23% of the queries will have cost of 3.

Total cost = \(\sum_{i} (\text{depth}(i) + 1) \times \text{freq}(i)\)

\[
= 1 \times 0.04 + 2 \times 0.4 + 2 \times 0.1 + 3 \times 0.05 + 3 \times 0.08 + 3 \times 0.1 + 3 \times 0.23 \\
= 2.42
\]
A different binary search tree (BST) leads to a different total cost:

Total cost = 1x0.4 + 2x0.05 + 2x0.23 + 3x0.1 + 4x0.08 + 4x0.1 + 5x0.04
= 2.18

This is in fact an optimal BST.
Optimal Binary Search Tree Problem

Given:

A collection of \( n \) keys \( K_1 < K_2 < \ldots K_n \) to be stored in a BST.

The corresponding \( p_i \) values for \( 1 \leq i \leq n \)

\( p_i \): probability of searching for key \( K_i \)

Find:

An optimal BST with minimum total cost:

\[
\text{Total cost} = \sum_{i} (\text{depth}(i)+1) \times \text{freq}(i)
\]

Note: The BST will be static. Only search operations will be performed. No insert, no delete, etc.
Cost of a Binary Search Tree

**Lemma 1**: Let $T_{ij}$ be a BST containing keys $K_i < K_{i+1} < \ldots < K_j$. Let $T_L$ and $T_R$ be the left and right subtrees of $T$. Then we have:

$$
cost(T_{ij}) = cost(T_L) + cost(T_R) + \sum_{h=i}^{j} p_h
$$

**Intuition**: When we add the root node, the depth of each node in $T_L$ and $T_R$ increases by 1. So, the cost of node $h$ increases by $p_h$. In addition, the cost of root node $r$ is $p_r$. That’s why, we have the last term at the end of the formula above.
Lemma 2: Optimal substructure property

Consider an optimal BST $T_{ij}$ for keys $K_i < K_{i+1} < \ldots < K_j$

Let $K_m$ be the key at the root of $T_{ij}$

Then:

- $T_{i,m-1}$ is an optimal BST for subproblem containing keys: $K_i < \ldots < K_{m-1}$
- $T_{m+1,j}$ is an optimal BST for subproblem containing keys: $K_{m+1} < \ldots < K_j$

$$\text{cost}(T_{ij}) = \text{cost}(T_{i,m-1}) + \text{cost}(T_{m+1,j}) + \sum_{h=i}^{j} p_h$$
Recursive Formulation

Note: We don’t know which root vertex leads to the minimum total cost. So, we need to try each vertex $m$, and choose the one with minimum total cost.

c[i, j]: cost of an optimal BST $T_{ij}$ for the subproblem $K_i < \ldots < K_j$

\[
c[i, j] = \begin{cases} 
0 & \text{if } i > j \\
\min_{i \leq r \leq j} \{c[i, r-1] + c[r+1, j] + P_{ij}\} & \text{otherwise}
\end{cases}
\]

where $P_{ij} = \begin{cases} 
p_h & \text{if } h = i 
\end{cases}$
Bottom-up computation

\[
c[i, j] = \begin{cases} 
0 & \text{if } i > j \\
\min_{i \leq r \leq j} \{c[i, r-1] + c[r+1, j] + P_{ij}\} & \text{otherwise}
\end{cases}
\]

How to choose the order in which we process \( c[i, j] \) values?

Before computing \( c[i, j] \), we have to make sure that the values for \( c[i, r-1] \) and \( c[r+1, j] \) have been computed for all \( r \).
\[
c[i, j] = \begin{cases} 
0 & \text{if } i > j \\
\min_{i \leq r \leq j} \{c[i, r - 1] + c[r + 1, j] + P_{ij}\} & \text{otherwise}
\end{cases}
\]
If the entries \( c[i,j] \) are computed in the shown order, then \( c[i,r-1] \) and \( c[r+1,j] \) values are guaranteed to be computed before \( c[i,j] \).
Computing the Optimal BST Cost

**COMPUTE-OPTIMAL-BST-COST** $(p, n)$

```plaintext
for $i \leftarrow 1$ to $n+1$ do
    $c[i, i-1] \leftarrow 0$

$PS[1] \leftarrow p[1]$  // $PS[i]$: prefix_sum$(i)$: Sum of all $p[j]$ values for $j \leq i$

for $i \leftarrow 2$ to $n$ do
    $PS[i] \leftarrow p[i] + PS[i-1]$  // compute the prefix sum

for $d \leftarrow 0$ to $n-1$ do
    for $i \leftarrow 1$ to $n - d$ do
        $j \leftarrow i + d$
        $c[i, j] \leftarrow \infty$
        for $r \leftarrow i$ to $j$ do
            $c[i, j] \leftarrow \min\{c[i, j], c[i, r-1] + c[r+1, j] + PS[j] - PS[i-1]\}$

return $c[1, n]$
```
Note on Prefix Sum

- We need $P_{ij}$ values for each $i, j$ ($1 \leq i \leq n$ and $1 \leq j \leq n$),

$$j$$

where: $P_{ij} = \sum_{h=i}^{j} p_h$

- If we compute the summation directly for every $(i, j)$ pair, the total runtime would be $\Theta(n^3)$.

- Instead, we spend $O(n)$ time in preprocessing to compute the prefix sum array $PS$. Then we can compute each $P_{ij}$ in $O(1)$ time using $PS$. 
Note on Prefix Sum

In preprocessing, compute for each $i$:

$$PS[i]: \text{the sum of } p[j] \text{ values for } 1 \leq j \leq i$$

Then, we can compute $P_{ij}$ in $O(1)$ time as follows:

$$P_{ij} = PS[i] - PS[j-1]$$

**Example:**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$:</td>
<td>0.05</td>
<td>0.02</td>
<td>0.06</td>
<td>0.07</td>
<td>0.20</td>
<td>0.05</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>$PS$:</td>
<td>0.05</td>
<td>0.07</td>
<td>0.13</td>
<td>0.20</td>
<td>0.40</td>
<td>0.45</td>
<td>0.53</td>
<td>0.55</td>
</tr>
</tbody>
</table>

$P_{27} = PS[7] - PS[1] = 0.53 - 0.05 = 0.48$

$P_{36} = PS[6] - PS[2] = 0.45 - 0.07 = 0.38$