Computer Science
Spring 2021 Lecture 16:
Dynamic Programming:
Edit Distance

PS3   Q1

ex   input (partial): $P_4 = 10\%$

$\qquad\qquad\qquad\qquad P_5 = 15\%$

$\qquad\qquad\qquad\qquad P_6 = 3\%$

$X[4,4] = 10\%$

$X[4,5] = 25\%$

$X[4,6] = 28\%$

$X[5,6] = 18\%$

Edit Distance

- ▶ Given two strings
- ▶ Convert first to second
- ▶ Insert, delete, or substitute characters

Example: convert FOOD to MONEY

FOOD → MOOD $\overset{sub}{\to}$ MOND → MONED $\overset{substitute}{\to}$ MONEY

Substitute   *insert E*

One way: align the words

free

F O O   D
M O N E Y

- ▶ The **cost** of an edit is # changes

2

Formulate this recursively

$X = \text{"}\begin{bmatrix} \downarrow \\ \ldots \end{bmatrix} j k\text{"}$

$Y = \text{"}\begin{bmatrix} \ldots & k \end{bmatrix} j\text{"}$

Define Edit(i, j) to be the **minimum cost** to convert $X[1 \ldots i]$ to $Y[1 \ldots j]$.
What happened in the last column?

▶ Insertion ($Y[j]$ was added) $\big(\text{or, } Y[j] \text{ was inserted from previous case}\big)$

$ins = 1 + Edit(i, j-1)$ *Recursive Substructure*

▶ Deletion ($X[i]$ was added) $\big(\text{or, } X[i] \text{ gets deleted to form previous case}\big)$

$del = 1 + Edit(i-1, j)$

▶ Substitution (or not)

$sub = Edit(i-1, j-1) + (X[i] \neq Y[j])$

Any special cases?

} *Overlapping Subproblems*

cost iff $X[i], Y[j]$ diff

if $i = 0$ return j

else if $j = 0$ return i

else   return $\min(ins, del, sub)$

Iterative

EditDistance($X[1 \ldots n], Y[1 \ldots m]$)
  for $j \leftarrow 0 \ldots m$ **do**   *declare* $Edit[0 \ldots n, 0 \ldots m]$
    $Edit[0, j] \leftarrow j$
  for $i \leftarrow 1 \ldots n$ **do**
    $Edit[i, 0] \leftarrow i$

    for $j \leftarrow 1 \ldots m$
      // compute $Edit[i, j]$
      $ins = 1 + Edit[i, j-1]$
      $del = 1 + Edit[i-1, j]$
      $sub = Edit[i-1, j-1] + (X[i] \neq Y[j])$
      $Edit[i, j] = \min(ins, del, sub)$

Filling in the Table

		A	L	G	O	R	I	T	H	M
	0	1	2	3	4	5	6	7	8	9
A	1	0	1	2	3	4	5	6	7	8
L	2	1		1	2	3	4	5	6	7
T	3	2	1	1						
R	4									
U	5									
I	6									
S	7									
T	8									
I	9									
C	10									

Reading the Table

		A	L	G	O	R	I	T	H	M
	0	1	2	3	4	5	6	7	8	9
A	1	0	1	2	3	4	5	6	7	8
L	2	1	0	1	2	3	4	5	6	7
T	3	2	1	1	2	3	4	4	5	6
R	4	3	2	2	2	2	3	4	5	6
U	5	4	3	3	3	3	3	4	5	6
I	6	5	4	4	4	4	3	4	5	6
S	7	6	5	5	5	5	4	4	5	6
T	8	7	6	6	6	6	5	4	5	6
I	9	8	7	7	7	7	6	5	5	6
C	10	9	8	8	8	8	7	6	6	6

I _ THM
I S T I C