Longest common subsequence (LCS):

Input: strings $X = x_1 x_2 \dots x_n$ & $Y = y_1 \dots y_n$

Goal: length of longest string Z which is a subsequence in both X & Y.

Example: $X = \underline{B}\,\underline{C}\,D\,\underline{B}\,C\,D\,\underline{A}$

$Y = A\,\underline{B}\,E\,\underline{C}\,\underline{B}\,\underline{A}\,A$

Solution $= 4$ corresponding to BCBA

Application: UNIX diff

Step 1: Define subproblem

Attempt 1: Consider prefixes.

For $0 \leq i \leq n$,

$L(i) =$ length of LCS in $x_1, \dots, x_i$ & $y_1, \dots, y_i$

Can we express $L(i)$ in terms of $L(1), \ldots, L(i-1)$

Earlier example: BCDBCDA
ABECBAA

For $L(7)$: $X_7 = Y_7$ so might as well match & take best of $L(6)$

So $L(7) = 1 + L(6)$

But what if $X_i \neq Y_i$? e.g., $L(6)$.

Might use X_i with Y_j for $j < i$

or Y_i with X_j

then left with dif't. length strings.

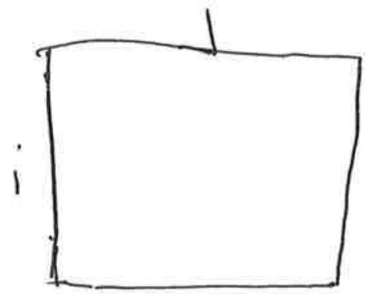
<u>Attempt 2</u>: For $0 \leq i, j \leq n$,

let $L(i,j)$ = length of LCS of $X_1, \ldots, X_i$ with $Y_1, \ldots, Y_j$

(can think of n replaced by n & m

length of X   length of Y

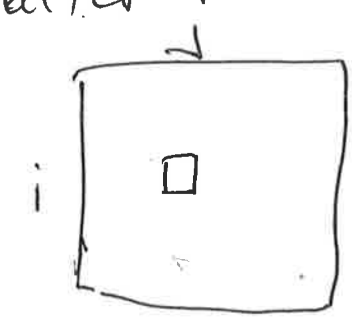
2-dimensional table $L =$



Base cases: 1st row & 1st column.

$$L(i,0) = 0 \quad \& \quad L(0,j) = 0$$

For $i > 1$, $j > 1$, let's figure out how to express $L(i,j)$ in terms of smaller



Two cases to consider:

① $X_i = Y_j$ or ② $X_i \neq Y_j$

Case ①: $X_i = Y_j$:

if LCS ends with $X_i = Y_j$ then: $L(i,j) = 1 + L(i-1, j-1)$.

What if X_i & Y_j not used then can match together & get longer CS.

If X_i matched with Y_ℓ for $\ell < j$ then can match with Y_j instead to get same length.

So, $L(i,j) = 1 + L(i-1, j-1)$.

2 Suppose $X_i \neq Y_j$:

Either X_i not matched &/or Y_j not matched (can't both be matched)

Hence, $L(i,j) = \max\{L(i-1,j), L(i,j-1)\}$

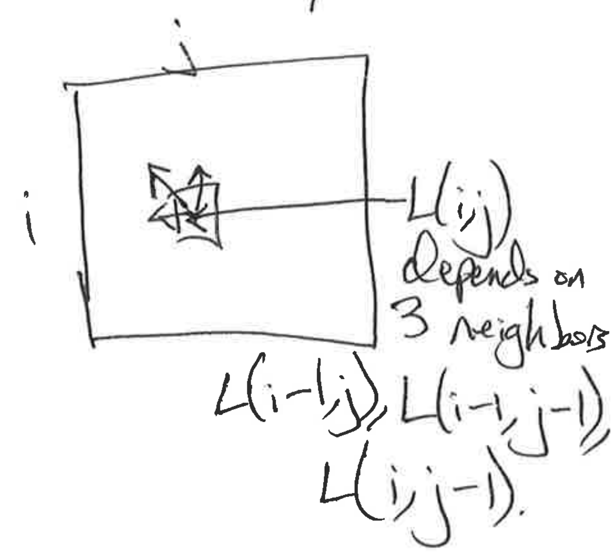
Summarizing:

$$L(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ 1+L(i-1,j-1) & \text{if } X_i = Y_j \\ \max\{L(i-1,j), L(i,j-1)\} & \text{if } X_i \neq Y_j \end{cases}$$

How to fill the table?

Row by row (or column by column)

L = 



LCS(X, Y): $X = x_1 \ldots x_n$, $Y = y_1 \ldots y_m$

For $i = 0 \to n$, $L(i,0) = 0$
For $j = 0 \to m$, $L(0,j) = 0$

For $i = 1 \to n$
    For $j = 1 \to m$
        if $x_i = y_j$
            then $L(i,j) = 1 + L(i-1, j-1)$
            else $L(i,j) = \max\{L(i-1,j), L(i,j-1)\}$

Return$(L(n,m))$.

Running time:

For inputs of size n & m,
one for loop of size $O(n)$
& inner loop of $O(m)$

$\Longrightarrow O(nm)$ total time.