Aashish Yadavally
Fall 2019

Homework IV

CSCI 6470
October 22, 2019

Question 1

Given two input sequences $x_1x_2x_3...x_n$ and $y_1y_2y_3...y_m$. Let $B(i,j)$ be the length of an *lcs* between suffixes $x_1x_2x_3...x_n$ and $y_1y_2y_3...y_m$.

$$B(i,j) = \begin{cases} 1 + B(i+1, j+1) & x_i = y_i \\ max(B(i+1,j), B(i,j+1)) & x_i \neq y_i \end{cases}$$

The base cases for $B(i,j)$ would be:
$B(n+1,k) = 0$ for $1 \leq k \leq m+1$ and $B(k, m+1) = 0$ for $1 \leq k \leq n+1$

Question 2

Given two sequences $x_1x_2x_3...x_n$ and $y_1y_2y_3...y_m$.

Given that the matching score of two set of indices $1 \leq i_i < i_2 < ... < i_k \leq n$ and $j_i < j_2 < ... < j_k \leq m$ for some $k \geq 0$ such that:

$$\Sigma_{r=1}^{k} \mu(x_{i_r}, y_{i_r}) = \mu(x_{i_1}, y_{i_1}) + \mu(x_{i_2}, y_{i_2}) + ... + \mu(x_{i_k}, y_{i_k})$$

Considering that the pairwise score is given by μ, which may be arbitrary, even if the two "tails" x_i and y_j are the same symbol, matching them may not necessarily lead to a globally optimal solution. Thus, the maximum matching score between prefixes $x_1x_2...x_a$ and $y_1y_2...y_b$ is given by:

$$S(i,j) = max \begin{cases} S(i-1, j-1) + \mu(x_{i-1}, y_{j-1}) \\ S(i, j-1) + \mu(x_i, y_{j-1}) \\ S(i-1, j) + \mu(x_{i-1}, y_j)) \end{cases}$$

The base cases for $S(i,j)$ would be:
$S(k,0) = 0$ for $0 \leq k \leq n$ and $S(0,k) = 0$ for $0 \leq k \leq m$

Question 3

Given an input sequence $x_1x_2...x_n$ consisting of four symbols '(', ')', '[', ']'.

The function $P(i,j)$ for $1 \leq i \leq j \leq n$ is the maximum number of paired parentheses in a legal expression as a subsequence $x_i...x_j$.

The maximum number of paired parentheses for an input sequence $x_1x_2...x_n$ is given by:

$$P(i,j) = max\begin{cases} P(i,k) + P(k+1,j) + Legality - Score(i,j) \\ P(i+1,k-1) + P(k+1,j) + Legality - Score(i,k) \\ P(i,k) + P(k+2,j-1) + Legality - Score(k+1,j) \end{cases}$$

In the first case, if (x_i, x_j) is legal, it translates to $P(i,j) = P(i,k) + P(k+1,j) + 1$. Otherwise, it translates to $max(P(i,j-1), P(i+1,j))$

We get the second case by computing $P(i,k)$ in the inside fashion, provided that (x_i, x_j) is legal, resulting in $P(i+1,k-1) + P(k+1,j) + 1$.

We get the third case by computing $P(k+1,j)$ in the inside fashion, provided that (x_{k+1}, x_j) is legal, resulting in $P(i,k) + P(k+2,j-1) + 1$.

Thus, $P(i,j)$ can be rewritten as:

$$P(i,j) = max\begin{cases} 1 + P(i+1,j-1), & \text{if } (x_i, x_j) \text{ is a legal pair, i.e, ("(",")") or ("[","]")} \\ \text{max\{P(i, j-1), P(i-1, j)\}} & \text{if } (x_i, x_j) \text{ is an illegal pair} \\ \text{max \{1 + P(i+1, k-1) + P(k+1, j), } if(\text{x}_i, x_k) \text{ is a legal pair;} \\ \qquad \text{1+ P(i, k) + P(k+2, j-1), } if(\text{x}_k, x_j) \text{ is a legal pair\}, for } i \leq k \leq j \end{cases}$$

The base case for $P(i,j)$ would be:
$P(i,j) = 0$, when $i \geq j$

Question 4

Proof:
Given two sequences $\mathbf{x} = x_1..x_m$ and $\mathbf{y} = y_1..y_n$.

Let the solution given by the algorithm for the longest common subsequence (LCS) be $z = z_1z_2...z_k$ where $1 \leq k \leq n$.

Since $x_m = y_n$, $z_k = x_m = y_n$.

Let us assume that there is a common subsequence z' for $x_1...x_{m-1}$ and $y_1...y_{n-1}$ such that the length of $z' > k - 1$.

If this is the case, the length of x_m or y_n appended to z' would be of length greater than k. However, the length of the longest common subsequence for $\mathbf{x}$ and $\mathbf{y}$ is k.

Thus, this is contradictory to our assumption that there is a common subsequence z' for x_{m-1} and y_{n-1} such that the length of $z' > k - 1$.

Thus, z' can have a maximum length of $(k - 1)$.

Thus, the common subsequence z' is an lcs for $x_1 ... x_{m-1}$ and $y_1 ... y_{n-1}$ with a length $(k - 1)$.

Consider an optimal subsequence for $x_1 ... x_{m-1}$ and $y_1 ... y_{n-1}$. Since z' is an lcs for these sequences, the optimal subsequence can be swapped with z'.

By extending this greedy solution to every optimal subsequence for two sequences, it can be concluded that a global optimum can be arrived at by selecting a local optimum.

Thus, $z = z'x_m$ is an lcs as well as an optimal subsequence $\mathbf{z}$ for $\mathbf{x}$ and $\mathbf{y}$, where z' is the lcs for $x_1 ... xm - 1$ and $y_1 ... y_{n-1}$

Therefore, this strategy does not compromise the optimality of the computed LCS.

Question 5

Let G be a gas station between city A and city B along US Route 1, where Michael just filled up his car tank.

Let us assume Michael can travel a maximum distance of d from A to a gas station H such that he does not run out of gas, and d is less than the distance between A and B.
Consider an optimal solution H_o, which is the optimal gas station that Michael should stop at after G. However, since H is the farthest from G such that Michael does not run out of gas, H_o should be closer to Michael than H.

Thus, H_o can be swapped with H in the optimal solution without affecting the optimality of the solution.

This can be extended to every greedy solution stop for Michael, thus replacing all the stops in the optimal solution with all the stops in the greedy solution, without affecting the optimality of the solution.

Therefore, the greedy solution can be established as the optimal solution, and the greedy solution cannot do worse than any optimal solution.

Hence proved that Michael's greedy strategy does allow him to stop at the fewest gas stations during his trip.