DSC 190 - Homework 06

Due: Wednesday, February 16

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. Unless otherwise noted by the problem's instructions, show your work or provide some justification for your answer. Homeworks are due via Gradescope at 11:59 PM.

Problem 1.

Consider the following set of activities.

Activity #	Start	Finish	Weight
0	0	3	12
1	1	2.5	10
2	2	3	22
3	2.75	4.25	14
4	4	8	33
5	4.5	11.5	11
6	5	6	12
7	5.5	6.5	7
8	8	10	5
9	9	12	19

a) Suppose the dynamic programming solution to the weighted activity selection problem is run on this data, and let `cache` be an array storing the memoized solution to each subproblem. In particular, `cache[i]` is the weight of the max weight schedule considering only the activities $i, i+1, \ldots, 9$. Note that we will consider two activities `x` and `y` to be compatible if `y.start` $\geq$ `x.finish` or `x.start` $\geq$ `y.finish`.

What are the entries of `cache`? You do not need to include the "dummy" entry of zero at the end of the array.

Hint: when checking your work, you should ask yourself: is it possible that `cache[i] < cache[i+1]`?

Solution:

```
[74, 74, 74, 52, 52, 31, 31, 26, 19, 19]
```

This can be obtained by essentially running the bottom-up algorithm by hand.

b) What is the weight of the optimal solution to the weighted activity selection problem for this data?

Hint: the right answer is greater than 70, less than 80.

Solution: 74

Programming Problem 1.

In lecture, we designed a top-down dynamic programming solution for the longest common subsequence problem. In this problem, you will create a bottom-up iterative solution.

In a file named `lcs_bup.py` create a function named `lcs_bup(a, b)` which accepts two strings, a and b, and returns the length of a longest common subsequence of a and b. Your code should take a bottom-up dynamic programming approach to solving the problem. Namely, it should be iterative and have $\Theta(mn)$

time complexity, where m and n are the lengths of a and b, respectively. The efficiency of your code will be ensured empirically, too, so make sure that your code isn't doing anything too inefficient (namely, that it isn't printing stuff, since that is slow).

There is starter code for this problem on the course page.

Hint: see the last discussion, as well as the lecture on the weighted scheduling problem, for strategies on converting a top-down DP solution to a bottom-up algorithm.

Solution:

```
def lcs_bup(text1, text2):
    """Find the length of the longest common substring between `a` and `b`.

    Your code should take a bottom-up dynamic programming approach. You may add
    additional optional arguments to this function or create helper functions.

    Parameters
    ----------
    a, b (str)
        The strings to be checked for common substrings.

    Returns
    -------
    int
        The length of the longest common substring between `a` and `b`. Can be zero.

    Example
    -------
    >>> lcs_bup("cal berkeley", "cow birkley")
    8
    >>> lcs_bup("uc san diego", "you see sandy eggo waffles")
    9

    """
    # BEGIN PROMPT
    dp = [0]*(len(text1) + 1) #the dp table assumes first char is empty string ""

    for j in range(len(text2)):
        dpn = [0]
        for i in range(len(text1)):
            if text1[i] == text2[j]:
                dpn.append(dp[i] + 1)
            else:
                dpn.append(max(dpn[i], dp[i + 1]))
        dp = dpn
    return dp[-1]
    # END PROMPT
```

Programming Problem 2.

In lecture, we designed a dynamic programming solution for the weighted activity scheduling problem. Our solution computed the *weight* of the best possible schedule, but did not return the schedule itself. Luckily, an optimal schedule can be computed using the `cache` array as long as we know for each event i, the *next* event j which begins on or after i's finish time.

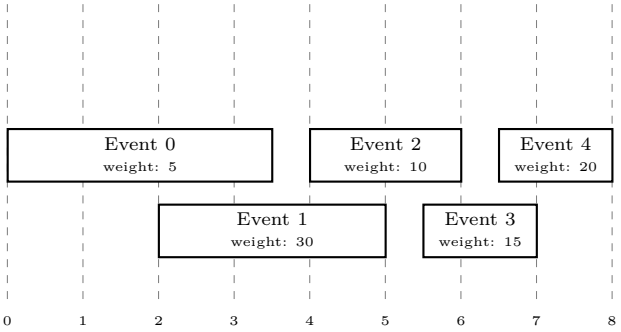
In a file named `recover_schedule.py`, write a function `recover_schedule(cache, next_activity)` which accepts two arguments:

- `cache`: The cache array as computed by the dynamic programming solution. It will be a Python list of numbers, where `cache[i]` is the weight of the best schedule that can be made from activities $i, i+1, \ldots, n-1$. `cache` *will* have one dummy entry of zero at the end, so that `len(cache)` is $n+1$, where n is the number of activities.
- `next_activity`: A list of n integers where `next_activity[i]` is the first activity starting after activity i finishes. We assume that the first activity is numbered zero, and that activities are sorted in order of start time. If there is no next activity possible, this entry should equal n.

Your function should return a list of integers representing the events that are in the optimal schedule. For instance, the list `[0, 2, 3]` represents the schedule consisting of the events 0, 2, and 3. There may be more than one optimal schedule; you need only return one of them.

Example: Consider the instance of the weighted scheduling problem shown below:

Activity #	Start	Finish	Weight
0	0	3.5	5
1	2	5	30
2	4	6	10
3	5.5	7	15
4	6.5	8	20



The best solution is to do activities 1 and 4, for a total weight of 50 (check for yourself!)

The dynamic programming `cache` will contain `[50, 50, 30, 20, 20, 0]`, where `cache[i]` represents the best solution using only activities from the set $\{i, i+1, \ldots, n-1\}$. The last entry of zero is the dummy entry mentioned above.

The `next_activity` list contains `[2, 3, 4, 5, 5]`. The 5s at the end denote that there is no possible next activity for Events 3 and 4.

Therefore, your code should behave as follows:

```
>>> cache = [50, 50, 30, 20, 20, 0]
>>> next_activity = [2, 3, 4, 5, 5]
>>> recover_schedule(cache, next_activity)
[1, 4]
```

Note that the code does not need to know anything about the actual activities, such as their weight – it only needs to know what the *next* activity is.

Hint: the hard part of this problem isn't writing the code – it is determining how to go from the `cache` to a schedule. Before writing any code, try to figure out the right strategy using a small example.

Solution: Let's say `cache[i] == cache[i + 1]`. What this means is that the weight of the best schedule possible using activities $\{i+1, \ldots, n-1\}$ is the same as the weight of the best schedule possible using events $\{i, \ldots, n-1\}$. In other words, including event i makes no difference. If event i were in the optimal schedule, there would be a difference.

Therefore, we loop through the cache. If we see `cache[i] > cache[i+1]`, then event `i` is in our optimal schedule. But we don't need to continue looping to `i + 1`, `i + 1`, etc. We know that the next event that can possibly be in our optimal schedule is `next_activity[i]`, so we skip ahead to it. In fact, if

we don't skip ahead, we're likely to add another event to the schedule that should not be added.

```
def recover_schedule(cache, next_activity):
    """Return the indices of the activities that should be selected.

    You may assume that the activities are labeled in increasing order of their
    start time. In the case that more than one schedule is optimal, return an
    arbitrary one.

    Parameters
    ----------
    cache : List[float]
        A list such that cache[i] is the largest total weight of any valid
        schedule that consists of elements from the set {i, i+1, i+2, ..., n-1}.
        If there are n activities, `cache` should have n + 1 elements, with the last
        element being a dummy element equal to zero.
    next_activity : List[int]
        A list of integers such that next_activity[i] is the index of the next
        activity whose start time is >= activity i's end time.

    Returns
    -------
    List[int]
        A list of the activities containing in an optimal schedule.

    Example
    -------
    >>> cache = [50, 50, 30, 20, 20, 0]
    >>> next_activity = [2, 3, 4, 5, 5]
    >>> recover_schedule(cache, next_activity)
    [1, 4]

    """
    # BEGIN PROMPT
    n = len(next_activity)
    schedule = []
    i = 0
    while i < n:
        if cache[i] > cache[i + 1]:
            schedule.append(i)
            i = next_activity[i]
        else:
            i += 1
    return schedule
    # END PROMPT
```