Ruben Tegnide HW3 Q.1

(1) (Finding elements near the median) (10 points) Given an unsorted array A of n distinct numbers and an integer k, where $1 \le k \le n$, design an algorithm that finds the k numbers in A that are closest in value to the median of A in $\Theta(n)$ time.

(Note: Finding the numbers that are closest in value to the median has no relationship in general with how these numbers are ordered in the unsorted array A.)

First we want to find the median. In order to do so we use the kth-smallcot algorithm where we pass in $\lfloor \frac{n+1}{2} \rfloor$ (where n is the length of the array) as k which will give us the median of the array.

Now that we know the median we can find the absolute difference between the median and every other element in the array. We will create an additional array that will hold arrays storing the absolute difference between the median and the clements value and the element value itself. Then we will iterate over the array and skip over the first element that matches our median. At every position we append the absolute difference between the median and the elements value as well as the elements value (for recovery porposes) within the newly created array

new array of arrays = [[|A.-A., A.], [|Az-A.|Az] ... [|As-A., As], [|A7-A., A7]]

Once we have this new array we can find the kth smallest difference by using the kth-smallest algorithm on the first element of every sub-array within our view array and we pass in our k from the original problem. This will give us the element with the kth smallest difference. With this element we partition the new array around this element and have elements with differences smaller than the pivot to go on its right and the others to go on its left.

[A'[i][i] = k | A[i][i] > k] La k+n smallest difference

Our newly partitioned array will have all differences less than the kth smallest difference on the left side of the array so now we can return all elements in A' at positions A'[I][Z] -A'[K][Z] and we will have all k values closest to the median.

We know this works because we first find the medium and then gtail the differences between each element and the medium. Then we find the kth smallest elements of the remaining elements and partition around that them veture the first k elements. The first k elements must contain the elements that are closest to the medium become their differences are minimized and any elements to the right of the first k elements will have larger differences. We also know there won't be any elements to the left of the first k elements because we partitioned around k so adding any more elements would surpass

our desired to

Time Analysis

closest
$$(A, k)$$
 $n = A$. long the median $= k$ th smallest $(\frac{n+1}{2})$ - finds median $= k$ th smallest $(\frac{n+1}{2})$ - finds median $= k$ th smallest $(\frac{n+1}{2})$ - finds median $= k$ th smallest $= k$ th $= k$

(2) (Finding quantiles) (20 points) For a set S of n numbers and an integer k, where $1 \le k \le n$, the kth-quantiles of S are k-1 elements from S whose ranks in S divide the sorted set into k groups that are of equal size (to within one unit).

Given an unsorted array A of n distinct numbers, design an algorithm that finds the kth-quantiles of A in $O(n \log k)$ time.

(Note: As an illustration, the 4-quantiles of a set of scores are the values that define the 25-, 50-, and 75-percentile cutoffs. Similarly, the 10-quantiles of a set are the values that define the 10-, 20-, 30-,, and 90-percentile cutoffs.)

We want to use a divide and conquer approach that will divide the problem into smaller gentiles we need to find until we are at the point where k=1. At this point we return nothing because a k-guartile will be the entire army as I group.

If k is greater then I we want to find the quantile separator that will be closest to the middle of the array. We will do this by finding which quantile will be closest to the middle without going over it $mid-q=\frac{L^2/2}{2}$

the suntile separator absort to mid will be $\lfloor \frac{4}{2} \rfloor = 2$

Next we want to find element that would be at the position mid-q if the array were sorted. So we use the kth smallest algorithm and pass in Limid-q. Wk. I This will find one of our quantile separates so we store it.

Using [mid-q. "/k] = [2.8/4]=4 we find we want to

find the 4th smallest element

Next we partition the array around the element we just found so all under smaller than it are on its left and larger values on its right.

Then, we want to recursively call the grantile finding algorithm but on the left half of the array up to the quantile we just found and on the right half going from the quantile of the end. In both the left and right case we prose the list of found quantiles as well as a new keyful to [1/2] for the right patien and Lk/2] for the left portion.

Within these recursive calls we will find all the quantiles that are to the right of the middle of the array if k > 1 or do nothing if k = 1 effectively finding all quantiles.

We know this works because find one quantile everytime we all the function and then divide the problem into subproblems. We work on the left side of the array with a k value equal to Lk/2] (which removes the quantile we just found to the left of the middle) and \(\frac{k}{2} \) for the right side to get all the guntiles on the right side of the middle. Dividing k in half as well as dividing the array in half easures we find all k-quantiles.

Time Analysis

Q (A, k, found)

if
$$k!=1$$
 do

 $n = A \cdot length$
 $mid-q := \lfloor k/z \rfloor$
 $val-g:= (k+h-smallest (A, \lfloor mid-q \cdot n/k \rfloor)) \} \Theta(n)$

found. append ($val-q$)

partition (A, $val-q$)

Q (ALI: $\lfloor mid-q \cdot n/k \rfloor$, $\lfloor k/z \rfloor$, found)

Q (A[$mid-q \cdot n/k \rfloor$, $\lfloor k/z \rfloor$, found)

 $T(k/z)$
 $T(n) = 2T(k/z) + Z\Theta(n)$
 $\log_z Z = 1$ $c = 1$ $d = 0$
 $\Rightarrow \Theta(n^c \log^{d+1} n) = n \log k$

Ruben Tegnida Q. 943

- (3) (Answering dynamic kth-smallest queries) (bonus) (10 points) Given a set of n numbers, the kth-smallest element can be found in Θ(n) time using the algorithm we learned in class. Suppose we have a dynamic set S of numbers that changes over time, and we want to be able to efficiently support the operations of
 - (a) inserting a new number into set S, and
 - (b) finding the kth-smallest element in the current set S, for any k.

We could support both operations by representing S as an unsorted array A. Inserting a new element would take just $\Theta(1)$ time. Finding the kth-smallest element, however, would take $\Theta(n)$ time. If there are many kth-smallest operations executed on set S, performing all of them will take a large amount of total time. We might like to speed up the time to find the kth-smallest, at a slight increase in the time to insert an element.

Design an algorithm that (a) supports inserting an element into S in $O(\log n)$ time, and (b) supports finding the kth-smallest element of S in $O(\log n)$ time as well. (Hint: Consider modifying a balanced binary search tree data structure.)

(4) (Longest increasing subsequence) (20 points) Given a string S of numbers, an increasing subsequence of S is any subsequence T of S such that the numbers in T, read left-to-right across T, are strictly increasing. For example, if S = (3, 1, 6, 2, 5, 4), an increasing subsequence of S is T = (1, 2, 4).

Design a dynamic programming algorithm that finds the longest increasing subsequence of a string S of length n in $\Theta(n^2)$ time.

(Note: Do not solve this problem by reducing it to the longest common subsequence problem. Instead design a dynamic programming algorithm from first principles.)

! There are two ways an optimal solution can begin select the last elementor not select:t

[Casel A[0:i-i] < A[0:i] where i is the length of the array we want to solve up to. If A[i] is larger than the largest value in the subproblem A[0:i-i] then we include A[i] 6/6

A · [*, *, * ... *, *] Subroblem A[o:i-1] A [i] is included in the solution ifitis greater in value than the greatest Selected value for A[o:i-17

Case Z A [::-] > A [o:i] so A [i] isn't part of the solution so we don't count it in this subarray

9/10 Z. Next we want to develop a recurrence equation for when ACI is part of the solution and when it is not. To do this we would recel to calculate the solutions to a subarray ending in i-1

> We would create a new array to hold the max runs for each subproblem with array length in where in = A, length

> > max-runs = new Array [n] , max-val =0,

Your structure max-runs [1,n]:= 0 (initialize all elements to 0) should not depends on evaluation order.

 $\max_{0 \le i \le n} \left\{ \begin{array}{l} \max_{1 \le i \le n} \left\{ \max_{1 \le i \le n} \left(\max_{1 \le i \le n} \left(\min_{1 \le i \le n} \left($

3. We want to compute the max-runs for each subproblem as i goes from 0 to a so we can see which pairs of numbers will increase the max-und total (inducting an increase in the longest increasing subsequence) or beep it the same (inducting no change to the longest increasing subsequence)

```
fill Max Puns (A, max-runs)
                      if A.length! = 0 do
                           for i:= 1 to n-1 de
                                  max-val = 0
                                  for j = 0 to i-1 do
                                         if A[i] < A[i] do
                                              max-val = max (max-val, AGI)
                                  end
                                  max_run[i] = max_val + 1
                            end
                           return max-runs
                     end
4. In order to recover the sequence of numbers that Broned the longest
            subsenquence of increasing numbers are first find the largest element in max-runs and record its index. Then, we will go from that Index to the left in the army to find the very next element that has
             a value of exactly I less than the max. Continue to do so until
             you hit the element with a value of 1.
              find-seq (A, max-runs) begin
                    cor-max := 0
                     n= max_runs. length
                     for i == n-1 down to 0 do
                          if max-runs [i] > wr-max
                                  [i] enu-xam = xam-nus
                                  max-index = i
                          end
                    end
                     seg = new Array
                     self. append ( Atmax index])
                     for X := cor-max-1 danto 1 do
                            max-index -= 1
                            for y:= max index down to 0 do
                                  if max-runo[y] == x
seq. append (A[y])
                                      break
                                 end
                          end
                                                    Une in order to return list in
                                                          increasing order rether then
                                                            decreasing order.
```

We know this works because we are starting with an aptimal solution where the next element in the sequence will be included in the solution or not. We then find the optimal solution for the minimum subproblem of an army with two elements and keeping track of how many numbers are smaller than AEIJ for every element of A. Then, we find the total number of increwing number sequences found for the reat element in A by using the solutions found for the previous subproblem. Once our new array holds all the total lengths of increasing numbers we can recover the original numbers that composed the longest subsequence by finding the max element in max-runs and then sping to the left and finding the next consecutive number less than the current max.

$$\sum_{1 \leq i \leq n-1} \Theta(i) = \Theta\left(\frac{n^{i+1}}{|+|} + \Theta(n^i)\right) = \Theta\left(\frac{n^2}{2} + \Theta(n^2)\right) = \Theta(n^2)$$

Ruber Tegnida FindSeq Q.4 for i := n-1 dounts 0 do if max_runs [i] > wr_max do wr_max = max_runs [i] max_index = i 0(1) \ \(\sigma \text{0}(1) = \text{O}(n-1+0+1) = \text{O}(n) \) end end for x := cur-max-1 to 1 do max-index -= 1 for y:= max_index down to 0 do if max-runo[y] == x do | seq. append (A[y]) break $\Theta(1)$ Σ $\Theta(1) = \Theta(\max-index - O+1)$ $OSYS max-index = \Theta(\max-index)$ end and $\sum_{1 \leq x \leq cur-max} \frac{\Theta(n)}{\log x - index} = \sum_{1 \leq x \leq n} \frac{\Theta(n)}{\log x} = \frac{1}{\log x} = \frac{1}$ Laif the army were in sorted order cur-mx would equal n so we can upper bound cur - max to n. In this cage max-index would be at n-1 so we can upper bound mex - index to - as well

 $=\Theta(n^2)$

T(n) = $\theta(n^2)$ + $\theta(n)$ + $\theta(n^2)$ =

La from fill MaxRons

Find Seq

(5) (Editing strings) (30 points) Given two strings A[1:m] and B[1:n], the edit distance between A and B is the minimum cost of a script that edits A into B. A script is a series of edit operations, each edit operation has a non-negative cost, and the cost of a script is the sum of the costs of its operations.

The allowed edit operations in a script are:

- · copy, which leaves a character unchanged, and has cost 0,
- substitute, which replaces a character a with another character b, and has cost c_{sub}
- insert, which adds a character a into a string, and has cost cins,
- delete, which removes a character a from a string, and has cost c_{del}, and
- transpose, which replaces two adjacent characters ab in a string by the characters ba, and has cost c_{tra}.

Design a dynamic programming algorithm to compute the edit distance between A and B and recover the corresponding edit script in $\Theta(mn)$ time. You may assume that an optimal script never edits a given character more than once. The costs of operations are part of the input to your algorithm.

(Hint: Order the operations in an edit script so they occur left-to-right across string A, and then examine the possible ways in which an optimal script could end.)

1) The optimal solution will consist of selecting between 5 options for each position in A. Casel An == Bn which will result in a copy A = A, Az Az ··· An-IAn

B = B, Bz Bz ··· Bn-IBn

Find optimel rolution
to subproblem

current subproblem

optimized by a copy
since An = Bn Case 2 An-1 == Bn will result in a transpose A= A, Az A3 ... An-1 An B= B, B2 B3 ... Bn-1Bn optimal solution to subproblem where An and An-1 are transposed Cure 3 An!= Bn and Any!= Bn subcase | |A| > |B| then delete A = A, A, A, A, ... An-1 An B = B, B, B, B, ... Bm-1 Bm

delete A.

2) Next we will develop a recurrence equation for filling out our table costs [IBI, IAI]. Our table will hold every letter of A in the columns and every letter of B in the rows. Then, we can compare every substring of A with every substring of B to see what the minimum cost of transforming A[I,n-I] to B[I,m-I] so we can use that to develop an optimal solution for the entire problem.

dist
$$(i, j-1) + C_{DEL}$$

dist $(i, j-1) + C_{DEL}$

dist $(i-1, j) + C_{ENS}$
 $(i-1, j-1) + C_{ENS$

3) In order to evaluate which combination of operations would result in the minimum cost we construct a 20 array that will hold the cost of previous subproblems in order to solve the current problem.

	// //	A	B
11 55	0	CDEL	Coel
B	CINS	Csus	min (Cook
C	CINS	min (sub)	min (Cocs) min (Cocs) (Cocs)

From the bottom-right corner we have the total cost of transforming Strong A into string B

41) Once we have the dist table filled out we can recover which operations produced the minimum cost by looking at which operation was chosen. Since we do not have actual values for the costs of our operations we are unable to calculate which operations would be specifically choosen. However, we can create an algorithm that tells us which path to take to recover.

3/3

If we hit an insert wego up a row dut (i-1,j)

If me hit a delete me so left a column dist (i,j-1)

If we hit copy

If we hit a substitute

I fue hit a translate

we go diagonal to the top-left dist (i-1,j-1)

Time Analysis
Evaluation Portlon:

In order to fill in a table with 181 rows and 1A1 columns one cell at a time we need to take O(181.1A1) time which would be O(nm).

Recovery Portlan:

while recovering we only do a number of compotetions equal to the number of operations done on A to get it to resemble B. The number of operations would be equal to the length of the largest string so this partwood take O (max (m,n)) time

$$T(n) = O(nm) + O(mex(m,n)),$$

La this will always be less than O(mn)

T(n)= O(mn)

Ruben Tequida 0.6

Then discuss relationship between

2 50/. 5

(6) (Discrete knapsack) (20 points) In the discrete knapsack problem, the input is a collection of n items with associated weights w_1, w_2, \dots, w_n and values v_1, v_2, \dots, v_n , and a capacity k. Item i has weight w_i and value v_i . All weights w_i and the capacity k are positive integers. The output is a subset S of the items $\{1, 2, \dots, n\}$, called a knapsack, such that the total weight of all the items in knapsack S is at most k, and the total value of all the items in S is maximum. In other words, a solution to the discrete knapsack problem is an optimal knapsack S of items that does not exceed the weight capacity k while having the greatest possible value.

Design a dynamic programming algorithm that solves the discrete knapsack problem in $\Theta(nk)$ time.

(Hint: Examine the items in the order $1, 2, \ldots, n$, and consider knapsacks of all possible capacities.)

optimal solution will consist of choosing 1 of Z options: first describe bon problem ends.

Then, remove last element in sol.

There will be a sol. to a subp.

4/6 casel: The last item is included in the solution If the lastitem A[n] is part of the solution then we must find the optimal solution to the subproblem with items A'[1: n-1] and a capacity k'= k - A[n], weight

> A [*, *, *, ..., *, *]
>
> Find the optimal in optimal solution solution withitem Alin-1] and a max expecity decreesed by the weight of the last item

Case Z: Item A[n] is not included in the optimal solution

A [*, *, *, ..., *, *] item will not be included in optimal solution solution withitem A[1:n-1] with the same max capacity as before

9 10/10 Z. Next we want to develop a recurrence equation for our table KS [n: k] which holds the max when for a given weight upto k and containing any possible items upto n.

To do this we need to calculate KS[n-1:k] which

would be the subproblem not including a and having a weight of k on k-A[n]. neight. $ks[i:j]:=\begin{cases} \max\left(\left(A[i], v_{\text{alive}} + ks[i-1:j-A[i], weigh+\right), ks[i-1:j]\right), & i \geq 1, j \geq 1\\ & \text{if } A[i], weigh+ \leq k\\ ks[i-1:j] & \text{if } A[i], weigh+ \geq k \end{cases}, & i \geq 1, j \geq 1\\ & \text{derivation} \end{cases}$ 3. We want to compute the maximum valve we can hold in our knapsack for every possible neight capacity up to le.

fill NS (A, k) begin n = A. length KS = new Array (n+1, k+1) for i = 0 to n+1 do KS[i:0] = 0 end j := 1 to k+1 do K 5[0: j] = 0 for i:= 1 to n+1 de Por j = 1 to k+1 do if A[i]. weight = k do KS[::j] = max ((KS[:-1:j-A[:]. weigh+]+A[:]. value), KS[:-1:j]) end else do Ks[i:j]= KS[i-1:j] and end end return KS end

This will create a table that holds the max value for every combination of items up to a desired i and for a max capacity of j.

NS: 0 1 2 ... k

for i = 2 j=2 we want to know if adding item 2 fits

the may weight requirement and If it does if

adding it will increase the value or be less than

another combination of items that fit the weight

compare NS[:-1:j] to A[i]. where + JUS [:-1:j-A[i]. weight

choose the larger of the two

2124. Once we have array KS completely filled out we know the max velve that the knapsack can hold will be at KS [n:k]. Since this gives us the value and not the items that give us the value we can work from the bottem-ny btof the table to the top-left to find which items gave us that max. value.

```
And Items (A, NS) begin

n:= NS. length

k:= NS[0]. length

cur_val:= NS[n:k]

items = []

while (n!= 0) begin

if NS[n-1:k] == cur_val do

n = n-1

end

else do

items. append (A[n])

n = n-1

k = k- A[n+1]. weight

end

return items
```

We know this works because we are using dynamic programming to calculate the max value we can hold in the knapsack for every possible weight from 0 - k and for all possible options of items from no items to all the items. When we are filling in the table we are computing the max value for a given weight awith a number of possible items so when we add another item we can subtract that it ems weight from our current weight and look at the items (not including the one we just added) that give us the max. value for the new weight. So the bottom-right cell of the table will give us the max value which we can work backwands from in order to get the items that gave us that value.

```
Time Analysis
```

For fill NS

for
$$i:=0$$
 to $n+1$ de

 $KS[i:0]:=0$

and

 $\theta(n+1)=\Theta(n)$

$$\int_{|\underline{z}|\leq k+1} \Theta(1) = \Theta(\sum_{|\underline{z}|\leq k+1}) = \Theta(k+1) + 1 + 1 - \Theta(k+1) = \Theta(k)$$

$$\sum_{1 \le i \le n+1} \Theta(k) = \Theta\left(\sum_{1 \le i \le n+1} k\right) = \Theta\left(k\sum_{1 \le i \le n+1}\right) = \Theta\left(k\left(n+1 \ne r\right)\right)$$

For find Items

while
$$(n!=0)$$
 begin

if $N3[n-1:k] == cur_vel de$
 $n=n-1$

end

else de

items, append $(A[n])$
 $n=n-1$
 $k=k-A[n+1]$ weight

end

end

$$T(n) = \Theta(n) + \Theta(nk) + \Theta(n+k) = \Theta(nk)$$