EC 504
Spring, 2021
HW 7

Due Friday, April 16, 8PM on Gradescope.

1. (15 pts) Consider the two words: *perplex* and *peruse*. Consider them as a sequence of letters. We would like to find the best alignment of the two letter sequences, with the following costs:

   - Incur a penalty of 1 for any skipped letter.
   - Incur a penalty of 3 for matching a consonant to a vowel.
   - Incur a penalty of 2 for mismatching a consonant to another consonant or a vowel to another vowel.
   - Incur a reward of 1 for correctly matching two letters.

   Using this reward structure, find the minimum cost alignment. State the cost of the alignment.

   Solution: Minimum cost = 1. Best alignment is shown below.

		p	e	r	p	l	e	x
	0	1	2	3	4	5	6	7
p	1	-1	0	1	2	3	4	5
e	2	0	-2	-1	0	1	2	3
r	3	1	-1	-3	-2	-1	0	1
u	4	2	0	-2	-1	0	1	2
s	5	3	1	-1	0	1	2	3
e	6	4	2	0	1	2	0	1

p	e	r	p	l			e	x
p	e	r			u	s	e	

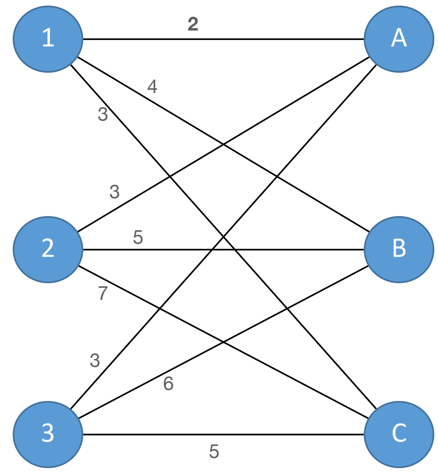
2. (20 pts) Consider the assignment problem illustrated in the figure on the right, where the costs are indicated on each of the edges. We are going to perform several steps of the successive shortest paths algorithm for solving this problem.

(a) As initialization, what are the initial prices of vertices A, B, C and vertices $1, 2, 3$?

(b) Let S be an added start node, and let T be an added destination node. Compute a shortest augmenting path consisting of vertices $S - 1 - A - T$, including the distance to all vertices. What are the adjusted prices of the vertices A, B, C and the vertices $1, 2, 3$ after computing the distances and modifying the prices?

(c) Perform the augmentation on path $S - 1 - A - T$. Compute and show the residual graph with the modified distances.

(d) Compute a new shortest augmenting path on this residual network, and compute the distances in the reduced graph to every node, and modify the prices for vertices $A, B, C, 1, 2, 3$.

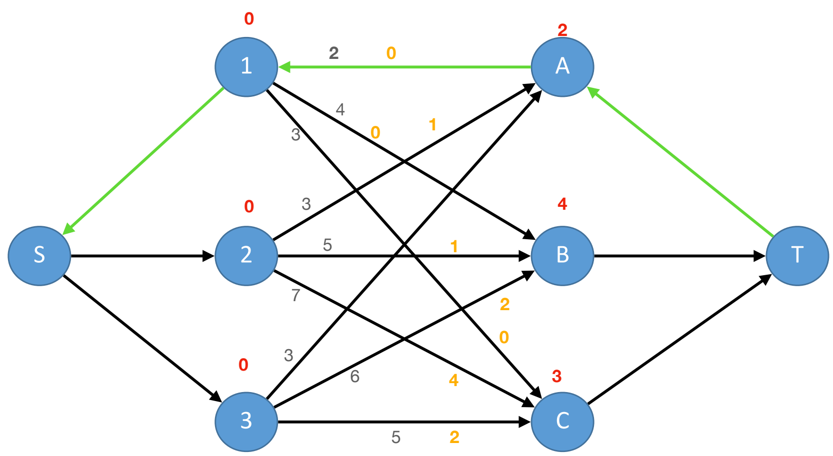


Solution:

(a) The prices of $1, 2, 3$ are 0. The initial prices are 2 for vertex A, 4 for vertex B, 3 for vertex C.

(b) The reduced cost distances to all the vertices in the shortest path computation are: 1, 2, 3, A, B, C are distance 0. Hence the prices remain 0 for $1, 2, 3$, and 2 for vertex A, 4 for vertex B, 3 for vertex C.

(c) The residual graph is shown below. Reverse edges are in green, vertex prices are in red, and reduced costs are in orange on each edge.



(d) The new shortest path will be S-2-B-T. The shortest distances from S in the residual network are 0 to 2, 3; 1 to B, T, A, C will be 1. The prices therefore will become: Price of 2, 3 remain at 0. Price of A rises to 3, price of 1 rises to 1. Price of B rises to 5, price of C rises to 4.

3. (10 pts) Consider the following pairs of coordinates: (7,2), (2,3), (4,6), (5,9), (8,1), (10,4), (13,5) (6,7), (9,8). Arrange them in a quadtree, using the maximum range 0-15 for each of the coordinates.

Solution: The boundaries for the quadtree areas will be $15/2^k$ for some integers k. Since the boundaries will not be integers, we won't have any points on the boundaries. The binary codes 00, 01, 10, 11 indicate which branch it is; 00 is the smallest x, smallest y (Southwest), 01 is smallest x, largest y (Northwest), 10 is largest x, smallest y (Southeast), and 11 is largest x and y (Northeast).

```
          O
          |
 ----------------- (7.5x7.5) region size
```

```
          |00   |01   |11    |10
          |     |     |      |
          |   (5,9) (9,8)    |
          |                  |
     ------------          -------------  (3.75x3.75) region size
     |00  |10    |11      |00    |01    |11
     |    |      |        |      |      |
   (2,3) (7,2)   |      (8,1)   (10,4) (13,5)
                 |
             ---------- (1.875x1.875)
              |01      |11
            (4,6)    (6,7)
```

4. (10 pts) Consider the following pairs of coordinates: (7,2), (2,3), (4,6), (5,9), (8,1), (10,4), (13,5) (6,7), (9,8). Show the 2-d binary search tree which results from inserting these elements in the order given, with no attempt to balance them.

Solution: Remember we alternate coordinates in the insertion comparisons. The root node is obvious.

```
          (7,2)
        /       \
    (2,3)       (8,1)
        \           \
      (4,6)        (10,4)
          \         /  \
          (5,9)   (9,8) (13,5)
          /
      (6,7)
```

5. (15 pts) Consider the following pairs of coordinates: (7,2), (2,3), (4,6), (5,9), (8,1), (10,4), (13,5) (6,7), (9,8). Form a balanced 2-d binary search tree using the median of the elements to split, with all the keys at the leaves of the tree. To make things unique, when you have to find a median of a list with an even number of entries, choose the smaller of the two entries as the median. Assume also that when a key is equal to the navigation key, it is added to the left subtree (that is, less than or equal).

Solution: Start by finding the median value of the first coordinate of all the points. Sorting by first coordinate yields the list:

$$(2,3),(4,6),(5,9),(6,7),(7,2),(8,1),(9,8),(10,4),(13,5).$$

The median value is 7, so we use 7 as a navigation key for the first split.

For the left subtree, the navigation key will be the median of the y coordinates of the entries that branch left. There are five entries: (2,3), (4,6), (5,9), (6,7), (7,2), so the median y coordinate is 6. For the right subtree, there are four entries: (8,1), (9,8), (10,4), (13,5), so there are two choices of median, so per the problem instructions, we pick the smaller, which is 4.

We repeat this process, alternating between branching on an x coordinate or a y coordinate

```
x-branch:               7
                     /             \
y-branch:        6                      4
              /      \              /       \
x-branch:    4         5          8           9
            / \      /   \      /   \       /   \
```

```
y-branch: 3 (7,2) (5,9) (6,7)(8,1)(10,4)(9,8)(13,5)
               / \
       (2,3)(4,6)
```

6. (20 pts) Consider the following pairs of coordinates: (7,2), (2,3), (4,6), (5,9), (8,1), (10,4), (13,5) (6,7), (9,8). Insert them as points into a 3-1 R tree, using increase in perimeter as the rule for navigation when selecting which leaf to insert. Consider a point as a rectangle with zero area and zero perimeter (i.e. the two corner points are the same.) Use the linear splitting rule to select seeds for splitting. Show the sequence of R trees that result from each insertion.

   Solution:

   Clearly, the first three insertions go directly into the root node; each entry in the root node consists of two points: bottom left and top right delineating a rectangle. The intial rectangles are a bit degenerate

```
 ____________________
|(7,2)| (2,3)| (4,6)|
|(7,2)| (2,3)| (4,6)|
|_____|______|______|
```

   Now, we add the fourth point and saturate the node; hence, we must split the node.

```
 ____________________
|(7,2)| (2,3)| (4,6)| (5,9)
|(7,2)| (2,3)| (4,6)| (5,9)
|_____|______|______|
```

   By the linear rule, the gap in x is 5, with a total span of 5, giving a ratio of 1. Similarly, the gap in y is 7, with a total span of 7, giving the same ratio. We pick the larger absolute gap of 7, so the seeds that are most separated in y are (7,2) and (5,9).

   We have two more rectangles to add. Consider first (2,3): If we add this to (7,2), we get an increase in area of 5 (d1((2,3))). If we add this to (5,9), we get an increase in area of 18 (d2((2,3))). The absolute value of this difference is 13.

   Consider next (4,6): Add to (7,2) gets an increase in area of 12 (d1((4,6))). Adding it to (5,9) gets an increase in area of 3 (d2((4,6))). The absolute difference is 9.

   Hence, the largest absolute difference comes from adding (2,3), so we add (2,3) to the seed (7,2) (the closest of the two seeds), forming the rectange with corners [(2,2),(7,3)] as the first group now, with area 5.

   We now add (4,6) to one of the groups. If we add it to the first group, the new enclosing rectangle would be [(2,2),(7,6)] with area 20, so the increase is 15. If we add it to (5,9), we get the rectangle [(4,6),(5,9)] with area 3. So clearly, we add it to (5,9). The result is to split the node into a new root node with navigation using the group rectangles, as:

```
                 ____________________
                |(2,2)| (4,6)|      |
                |(7,3)| (5,9)|      |
                |_____|______|______|
               ____|      |_________
               |                    |
     __________|______     _________|_______
    |(7,2)| (2,3)|    |   |(4,6)| (5,9)|     |
    |(7,2)| (2,3)|    |   |(4,6)| (5,9)|     |
```

We now insert (8,1). The perimeter of the first rectangle is 12. Adding (8,1) makes the rectangle [(2,1),(8,3)] with perimeter 16, leading to an increase of 4. the perimeter of the second rectangle is 8. Adding (8,1) to it makes the rectangle [(4,1),(8,9)] with perimeter 26. So we navigate to the first rectangle and insert it there (there is room). We change the navigation rectangle at the top to include this point.

```
              _____________________
              |(2,1)| (4,6)|      |
              |(8,3)| (5,9)|      |
              |_____|______|______|
             ____|      |_________
             |                    |
   __________|_______     _________|________
  |(7,2)| (2,3)|(8,1)|    |(4,6)| (5,9)|      |
  |(7,2)| (2,3)|(8,1)|    |(4,6)| (5,9)|      |
```

Next we add (10,4). The first rectangle in the root has perimeter 16, and would grow to [(2,1),(10,4)] with perimeter 22. The second rectangle has perimeter 8, and would grow to [(4,4),(10,9)] with perimeter 22. So we add again to the first rectangle, which will again saturate the node. We compute the linear split below:

```
  |(7,2)| (2,3)|(8,1)|(10,4)|
  |(7,2)| (2,3)|(8,1)|(10,4)|
```

This time the largest spread is in x, and there is a tie in normalized spread, so we choose the x coordinate to split. The two seeds will be (2,3) and (10,4). As before, we compute the change in area for adding (7,2) to group 1 (2,3), as 5 (d1), and to group 2 (10,4) as 6, for an absolute difference of 3. Adding (8,1) to group 1 changes area by 12, and to group 2 changes area by 6, for an absolute difference of 6. Hence, we add (8,1) to group 2 (largest absolute difference, nearest group) to form the new group 2 with rectangle [(8,1),(10,4)].

We compute adding (7,2) to this new group to give a new rectangle [(7,1),(10,4)] for a net increase in area of 6. Adding to the old group 1 increased the area by 5, so we add it to group 1 to form a new group with rectangle [(2,2),(7,3)].

We now split the nodes. we update the navigation rectagles at the root node, and obtain the following R-tree:

```
                     ___________________
                    |(2,2)| (4,6)|(8,1) |
                    |(7,3)| (5,9)|(10,4)|
                    |_____|______|______|
             __________|      |       |_______________
             |                |___                   |
   __________|_______     _________|________     ______|___________
  |(7,2)| (2,3)|     |    |(4,6)| (5,9)|     |   |(8,1)|(10,4)|      |
  |(7,2)| (2,3)|     |    |(4,6)| (5,9)|     |   |(8,1)|(10,4)|      |
```

We add (13,5). Increase in perimeter for the left rectangle is 16. Increase for the second rectangle is 18. Increase for the 3rd rectangle is 10. Hence, we add it to the 3rd rectangle, and modify the navigation rectangle to get:

```
                     ___________________
                    |(2,2)| (4,6)|(8,1) |
                    |(7,3)| (5,9)|(13,5)|
                    |_____|______|______|
             __________|      |       |_______________
             |                |___                   |
```

```
  __________|_______     _________|________     ______|____________
 |(7,2)| (2,3)|     |   |(4,6)| (5,9)|     |   |(8,1)|(10,4)|(13,5)|
 |(7,2)| (2,3)|     |   |(4,6)| (5,9)|     |   |(8,1)|(10,4)|(13,5)|
```

We add (6,7). Increase in perimeter for first rectangle is 10. Increase in perimeter for 2nd rectangle is 2. Increase for 3rd is 4. So we add it to 2nd rectangle leaf (there is room), and update navigation to get:

```
                     __________________
                    |(2,2)| (4,6)|(8,1) |
                    |(7,3)| (6,9)|(13,5)|
                    |_____|______|______|
             __________|      |      |_______________
             |                |___                  |
  __________|_______     _________|_________    _____|_____________
 |(7,2)| (2,3)|     |   |(4,6)| (5,9)|(6,7)|  |(8,1)|(10,4)|(13,5)|
 |(7,2)| (2,3)|     |   |(4,6)| (5,9)|(6,7)|  |(8,1)|(10,4)|(13,5)|
```

Last, we add (9,8). Change in perimeter for first rectangle is 14. For second rectangle, it is 6. For 3rd rectangle, it is also 6. Hence, we have a tie. We add it to the rectangle with the smallest area, which is Rectangle 2 at the root. This will saturate the leaf, which we will have to split. The leaf is:

```
                  ________________________
                 |(4,6)| (5,9)|(6,7)|(9,8)|
                 |(4,6)| (5,9)|(6,7)|(9,8)}
```

As before, we pick seeds by spread, and we find the x coordinate to be the farthest spread. The two seeds are (4,6) and (9,8). Adding (5,9) to group 1 increases area to 3 (d1), adding to group 2 increases area to 4, so absolute difference is 1. Adding (6,7) to group 1 increases area by 2, and to group 2 increases area by 3. Again the absolute difference is 1. With ties, choose the one that increases the area the least, so add (6,7) to group 1, increasing group 1 to [(4,6),(6,7)]. Now, adding (5,9) to this increases area by 4, and adding (5,9) to group 2 increases area by 4. Tie again goes to the group with the smallest area, which is group 2, to form rectangle [(5,8),(9,9)]. We get the following R-tree as a result of the split:

```
                     _________________________
                    |(2,2)| (4,6)|(5,8)| (8,1)|
                    |(7,3)| (6,7)|(9,9)|(13,5)|
                    |_____|______|_____|______|
             __________|      |      |      |____________________________
             |                |      |__________                        |
  __________|_____     _________|_______    ______|_________       _________|__________
 |(7,2)|(2,3)|    | |(4,6)|(6,7)|     | |(5,9)|(9,8)|   |     |(8,1)|(10,4)|(13,5)|
 |(7,2)|(2,3)|    | |(4,6)|(6,7)|     | |(5,9)|(9,8)|   |     |(8,1)|(10,4)|(13,5)|
```

Note that we have now saturated the root, so we must split it! We show this below:

```
                     _________________________
                    |(2,2)| (4,6)|(5,8)| (8,1)|
                    |(7,3)| (6,7)|(9,9)|(13,5)|
                    |_____|______|_____|______|
```

To find seeds, note that the largest x_{low} is 8, and the smallest x_{hi} is 6, for a spread of 2, and the range of x is 11, for a normalized ratio of 2/11. On the y axis, the smallest y_{hi} is 3, and the largest y_{low} is 8, for a spread of 5, and the toal range of y is 8. Hence, we select seeds as separate in the y direction, corresponding to the rectangles [(5,8),(9,9)] and [(2,2),(7,3)]. Adding rectangle [(4,6),(6,7)] to group 1 increases area by 11 (d1);

adding to group 2 increases area by 15 (d2), for an absolute difference of 4. Adding rectangle [(8,1),(13,5)] to group 1 increases the area by 60 (d1), and to group 2 increases the area by 39. Hence, the largest absolute difference is for rectangle [(8,1),(13,5)] added to group 2, to increase group 2 to [(2,1),(13,5)]. Now, adding [(4,6),(6,7)] to group 2 increases the area by 26, so it is best to add this to group 1 to form the rectangle [(4,6),(9,9)]. The resulting R-tree is shown below:

```
                        ____________________
                       | (4,6)|(2,1) |      |
                       | (9,9)|(13,5)|      |
                       |______|______|______|
                           |      |____________
                  _________|_______    ________|_________
                 |(5,8)|(4,6)|     |  |(2,2)|(8,1) |     |
                 |(9,9)|(6,9)|     |  |(7,3)|(13,5)|     |
                 |_____|_____|_____|  |_____|______|_____|
         __________|      |             |     |___________________
         |                |             |_                        |
 __________|_____   ________|_______   ______|_________    ________|__________
|(5,9)|(9,8)|    | |(4,6)|(6,7)|    | |(7,2)|(2,3)|   |    |(8,1)|(10,4)|(13,5)|
|(5,9)|(9,8)|    | |(4,6)|(6,7)|    | |(5,9)|(9,8)|   |    |(8,1)|(10,4)|(13,5)|
```