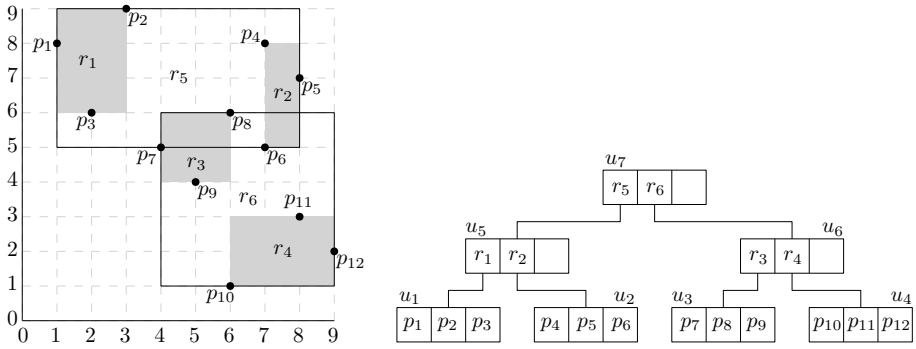
INFS 4205/7205: Exercise Set 4

Prepared by Yufei Tao and Junhao Gan

All the following problems will be based on the dataset and R-tree below:



Problem 1. What is the skyline of the dataset?

Solution. $p_1, p_3, p_7, p_9, p_{10}$.

Problem 2. Suppose that we use the NN algorithm to find the first skyline point, after which the NN algorithm will remove some points from the dataset. What are the *remaining* points in the dataset?

Solution. p_1, p_2, p_3, p_7, p_9.

Problem 3. Indicate the nodes of the R-tree that are accessed by the BBS algorithm to find the skyline.

Solution. $u_7, u_6, u_5, u_3, u_4, u_1$.

Problem 4. We know that every point in the skyline minimizes some monotonically increasing function. Give a function that is minimized by point $p_7 = (4, 5)$.

Solution. One such function is $f(x, y) = (\frac{x}{4})^{10} + (\frac{y}{5})^{10}$.

Problem 5. Consider the top-1 query that seeks the point minimizing $f(x, y) = x + 2y$. Explain how to convert this into nearest neighbor search. Specifically:

- What is the query point (of the nearest neighbor search) and what is the distance function?
- What are the nodes that are accessed by the best first algorithm in order to perform the nearest neighbor search?

Solution. The query point is $q = (0, 0)$ and the distance function between q and a point $p = (x, y)$ is $dist(q, p) = x + 2y$.

The best algorithm accesses u_7, u_6, u_4 (p_{10} is the result for the top-1 query).

Problem 6. Consider the BBS algorithm again. Recall that in the sorted list H, it orders the entries by their mindists from the origin. In the class, we calculated the mindist of a rectangle r

as the Euclidean distance from its bottom-left corner (x, y) to the origin (assuming 2D), namely, $\sqrt{x^2 + y^2}$. Now, we change the mindist definition to the L_1 distance (from the bottom-left corner to the origin), namely, $x+y$. Run the BBS algorithm again on the R-tree shown above, and indicate the nodes accessed.

Solution. $u_7, u_6, u_5, u_3, u_4, u_1$.

Problem 7. Prove that the BBS algorithm accesses precisely the same nodes no matter whether mindist is defined in Euclidean distance or L_1 distance.

Proof. The algorithm accesses only the nodes whose MBRs intersect the search region as defined in the class—the proof in the slides holds verbatim for L_1 distance as well.