1   Space filling curves for NN-queries   (1 P.)

1. Implement space-filling z-curves in a language of your choice. Your program has to take two files which contain points as input. The first file contains all the points of the base data set, while the second file contains query points. The program has to calculate and display:

   - The k-NN, of each query point, in the base data set, based on the actual distance.
   - The k-NN, of each query point, in the base data set, based on the z-curve distance.

   You can use the template in OLAT, which already parses the files and provides utility classes. Submit the code and the output of your program when executed with $k = 3$ and the two data files provided in OLAT. If you do not use the template, also submit instructions on how to compile and execute your program.
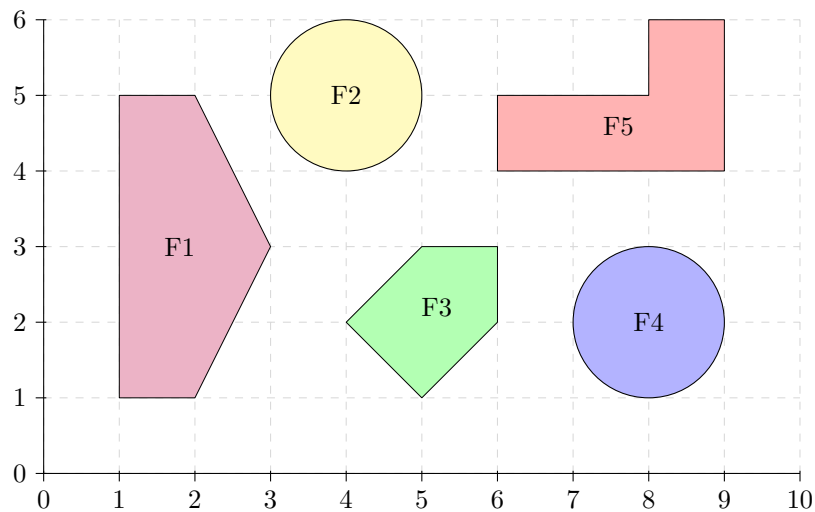
   Note: If you use code from external sources, provide the source as a comment.

2. Which differences can you see between the two results of your implementation? Explain why are or why are not the results the same.

2 R Tree (1 P.)

Given the following two-dimensional objects:

For the following R tree operations, explain exactly which steps are performed:

1. *Store the objects F1, F2, F3, F4, F5 in an initially empty R tree.* One node fits 1–2 entries.

2. *Find all objects, containing the point (6,2).*

3. *Find all objects that are positioned completely in the rectangle Q, which is defied by the points* $(2, 1)(9, 3)$.

4. *Find all objects, intersecting with Q': $(2, 2)(4, 4)$.*

   We assume that sharing exactly one point, also counts as intersecting.

3 Index Structures in Metric Space (1 P.)

Show, for each of the following distance functions, that the properties for being a metric are fulfilled, or provide a counterexample.

1. For two vectors (or points) $a = (x_a, y_a)$ and $b = (x_b, y_b)$:

   $d(a, b) \mapsto |x_a - x_b| + |y_a - y_b|$.

2. For two vectors (or points) $a = (x_a, y_a)$ and $b = (x_b, y_b)$: $d(a, b) \mapsto (x_a - x_b)^2 + (y_a - y_b)^2$.

3. For two strings s_1 and s_2, where S is the set of characters of the string (e.g.: $S(\text{"Codd"}) = \{\text{'C', 'o', 'd'}\}$): $d(s_1, s_2) \mapsto 2 * |S(s_1) \cap S(s_2)| / (|S(s_1)| + |S(s_2)|)$.
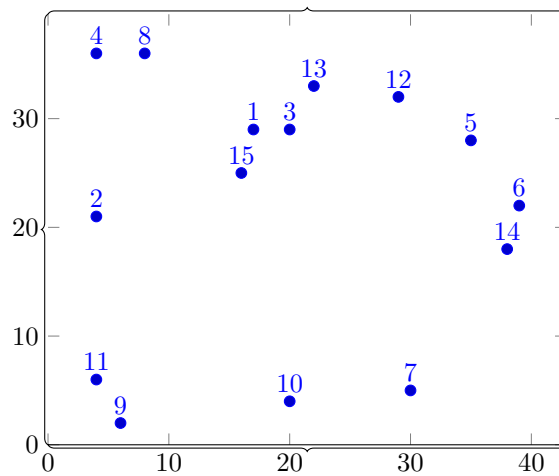
4 Misc Metric Indexing (1 P.)

Below are the points used for the following questions:

$$1 : (17, 29) \quad 2 : (4, 21) \quad 3 : (20, 29) \quad 4 : (4, 36) \quad 5 : (35, 28)$$

$$6 : (39, 22) \quad 7 : (30, 5) \quad 8 : (8, 36) \quad 9 : (6, 2) \quad 10 : (20, 4)$$

$$11 : (4, 6) \quad 12 : (29, 32) \quad 13 : (22, 33) \quad 14 : (38, 18) \quad 15 : (16, 25)$$

1. **GH Tree partitioning:** Create a GH partitioning, such that the leaf nodes of the tree have at most 2 elements. Draw the tree and draw the partitioning into the plot below.

2. **VP Tree:** Given the following VP tree, search for query point (25,4) with $\varepsilon = 11$. Describe which parts of the tree you pruned and why.