# Exam of Informatics - R Programming

We have a data frame `points` with two columns $x$ and $y$ representing points in a two dimensional space. For each point $p_i$ we want to find the point $p_{nn(i)}$ such that his distance $d(p_i, p_{nn(i)})$ is the minimum with respect to the distances $d(p_i, p_j), \forall j \notin \{i, nn(i)\}$ of the other points

| | x | y | nn | nn_dist |
|---|---|---|---|---|
| 1 | 1 | 1 | 7 | 0.58309518948453 |
| 2 | 2 | 1.5 | 7 | 0.53851648071345 |
| 3 | 4 | 4 | 4 | 1.4142135623731 |
| 4 | 3 | 3 | 3 | 1.4142135623731 |
| 5 | 1.5 | 3 | 4 | 1.5 |
| 6 | 4 | 1.5 | 4 | 1.80277563773199 |
| 7 | 1.5 | 1.3 | 2 | 0.53851648071345 |

The solution must not assume that the x values are sorted. Write a function which takes as parameters a numeric vector x and a positive integer n and returns the vector of discretised values xd.

Translate in R the pseudo-code algorithm(s) provided in next page.

Continue

# Algorithm

```
Given a dataframe whose columns represent the coordinates of points
in a d-dimensional space, for each point find the index and the distance
of the nearest neighbour
Method:
for each point scan the entire dataset to find the other point
with minimal distance
Parameters:
- df: a dataframe where each row represent a point in a d-dimensional space
Uses:
- i:  index for the main loop on points
- j:  index for the secondary loop on the neighbours of a point i
- dist_nn_i:  distance of the nearest neighbour of i
- nn_i:       nearest neighbour of i
- dist_ij:  distance of the current pair i,j
Algorithm:
- prepare an empty dataframe with two columns: the index of the
  nearest neighbour and its distance from the corresponding point
- repeat for each row i of the dataframe
  - initialize the minimum distance dist_nn from i to
    the maximum double
  - for each row j in the dataframe different from i
    - compute the distance dist_ij between the point in i and the point in j
    - if dist_ij is smaller than the current minimum dist_nn_i
      - set dist_nn_i to dist_ij
      - set nn_i to j
  - store the pair nn_i,dist_nn_i to the i row of the output
- return the output dataframe

#Test program
- read the file into a dataframe df
- store in a dataframe  df\_nn the result of the call of the function with parameter df
- print df\_nn
```