Deadline:

- Group 1: Monday 01.05 - 09:00 (!)
- Group 2: Wednesday 03.05 - 13:00

You can submit your solution via LearnWeb in the respective Exercise item. Please **only** submit in your respective group that you are attending, not in both.

1 Load the Data

1. Download the provided data file `cluster_data.csv`
2. Load the contents into your program:
   Recommended way is `pandas` → `pd.read_csv("cluster_data.csv", header=0, index_col=0)`
3. Plot the data with `matplotlib` (See Scatter-plot for examples)

2 Ramp Up

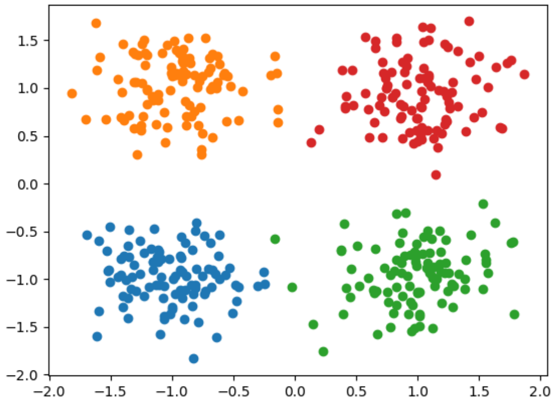
As a preparation for the full K-Means algorithm you can implement the distributed center computation

Create a program that can compute the centroid of the entire dataset in a distributed way by implementing step 5. and 6. from slide 10

Test your program with any amount of workers by distributing the data evenly among them
The result needs to be a single [x, y] coordinate close to [0, 0]

3 K-Means

Implement the K-Means algorithm as an MPI program.
Distribute the data from `cluster_data.scv` evenly across the ranks and follow the steps from slides 8-10
Test your program with $k = 4$ and any amount of workers you wish
Count the number of assigned points per cluster and (optional) visualize the points with color-coded cluster assignments.



Example of a color-coded scatter plot of `cluster_data.scv`