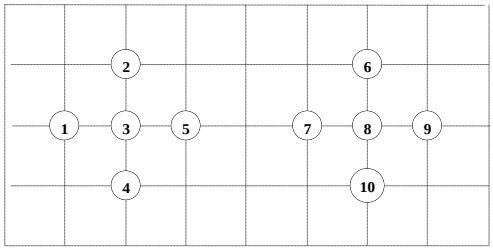
Clustering Methods

Exercises 7/7, 8.3.2022

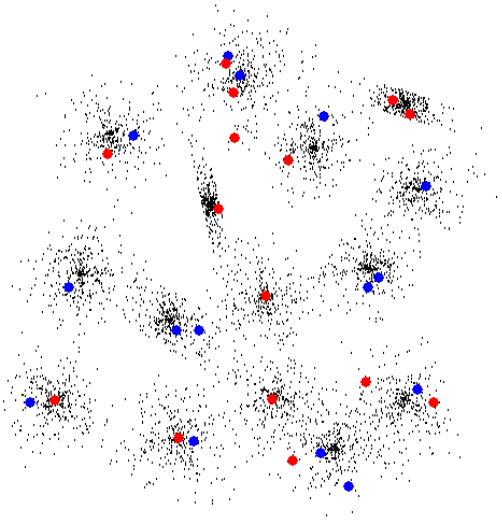
Five tasks required. If done more, each earn one bonus point.

1. Implement divisive clustering algorithm. Make it as simple as possible. Cluster S1-S4 datasets and calculate SSE (or CI) values and report processing time. The one finding the correct result to all four

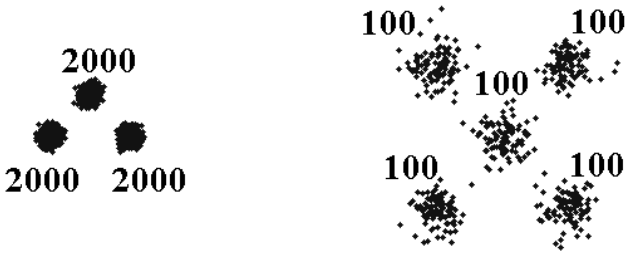
2. Show step by step how the divisive clustering algorithm works for the following dataset.



3. In genetic algorithm it would be possible combine not only two parents but several. We could even take centroids from all solutions and find the selected centroids from there. Invent simple algorithm to do this. Demonstrate using the dataset below. The red and blue points are examples of two parents.



4. Would the crossover algorithm from task 3 work for the unbalance dataset?



5. Can we use outlier detection in GA to detect which parent centroid to select? If yes, show how. If not, explain why not.