**Fundamentals of data science**  
Outlier detection – solutions

1. Draw three dimensional (so called group scatter plot) with classes indicated by different color of data points.



1. Please use one dimensional outlier detection using typical mean and standard deviation (3-sigma) method for all attributes separately. Note the number of outliers as **p**.

Five outliers (and features for which the outlier was found

x10 , x63 , x175 , x216, x253 (third attribute)

1. Please find outliers using distance-based multivariate method, with k=5, **p** found in point II. Namely: outliers are **p** cases for which distance to the 5th nearest neighbor is the largest.

We find 5 nearest neighbors for each point and select 5 points for which the distance to the 5-th nearest neighbors is the biggest:

x63, x175 , x10 , x216, x306

and distances:

29.13, 22.38, 15.42, 13.07, 10.48

1. Are the results for both methods the same?

In four out of five cases – yes!