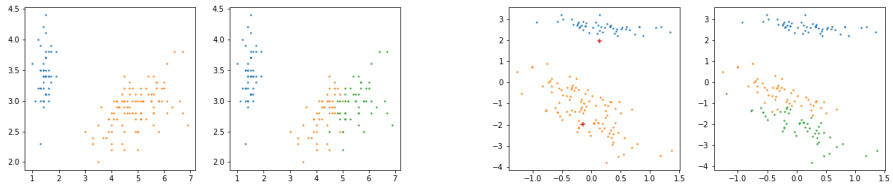
Homework 6 report

By Fedor Ivachev 费杰 2019280373

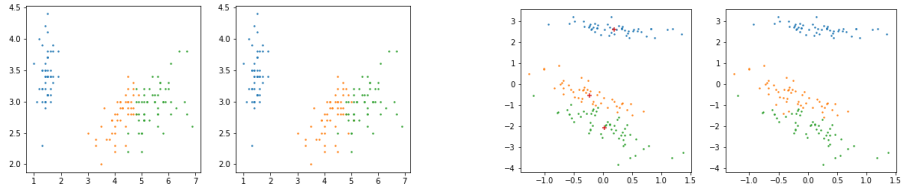
The task was to implement the EM algorithm for Gaussian mixture models as described in class.

The visualized plots are (predicted, true):

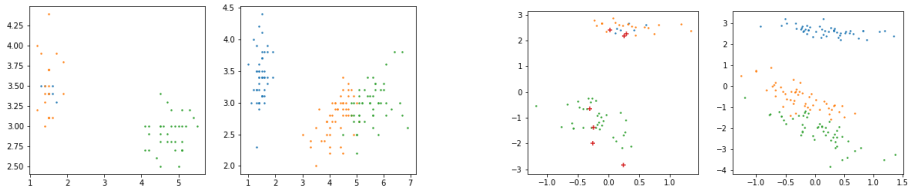
2 clusters (Cross, PCA):



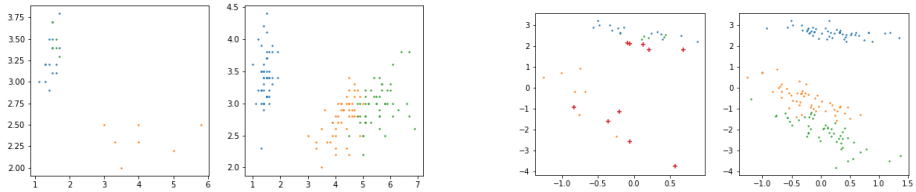
3 clusters (Cross, PCA):



7 clusters (Cross, PCA):



10 clusters (Cross, PCA):



The algorithm worked the best when number of clusters was equal to the number of species.

If the number of clusters is two, then the points from close clusters become one cluster. With the increasing the number of clusters, we see more and more incorrected predictions (because of increasing number of extra not needed clusters).

Also, the first step of EM-algorithm means a lot. The output really depends on the division of the dataset (which is performed randomly).