# Algorithmic Data Analysis

Esther Galbrun

Spring 2024

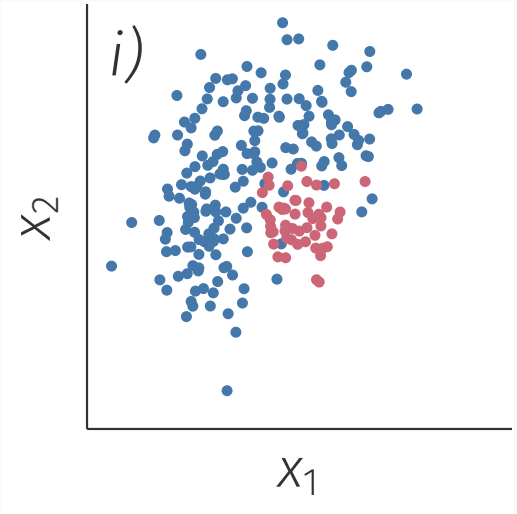
UNIVERSITY OF
EASTERN FINLAND

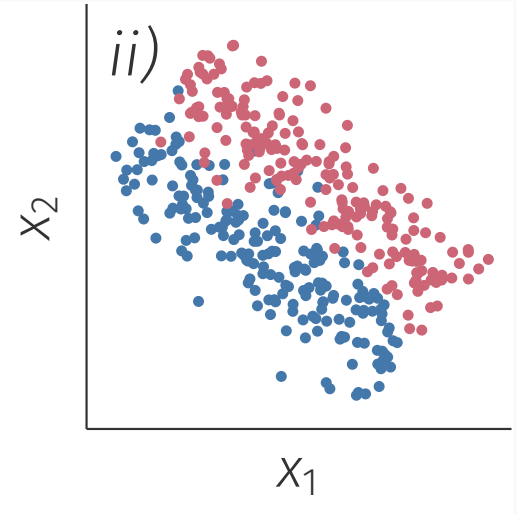
*Last updated on 2023-12-28 11:36*

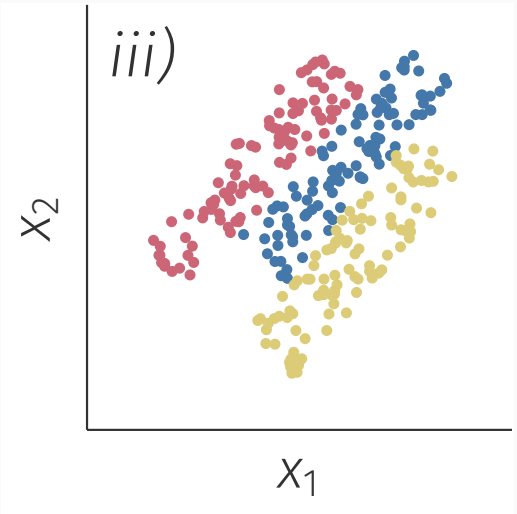
# Q2.1: Classification problems (i)

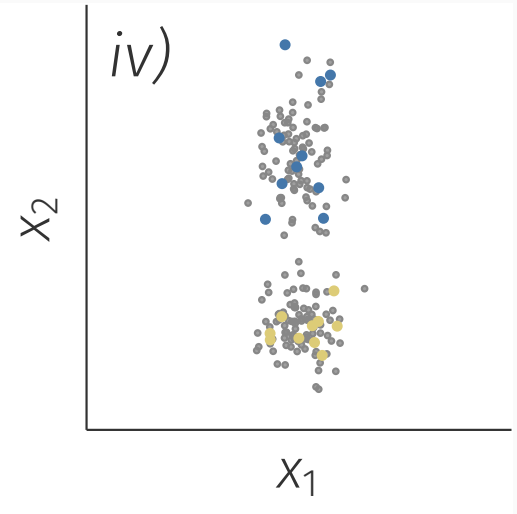
Associate each dataset with suitable keywords.

EM-algorithm
hard-margin SVM
kernel SVM
linearly separable
multi-class learning
one-against-one
oversampling
rare-class learning
reweighting
semi-supervised learning
soft-margin SVM
unbalanced data



# Q2.2: Algorithm(s) of choice (i)

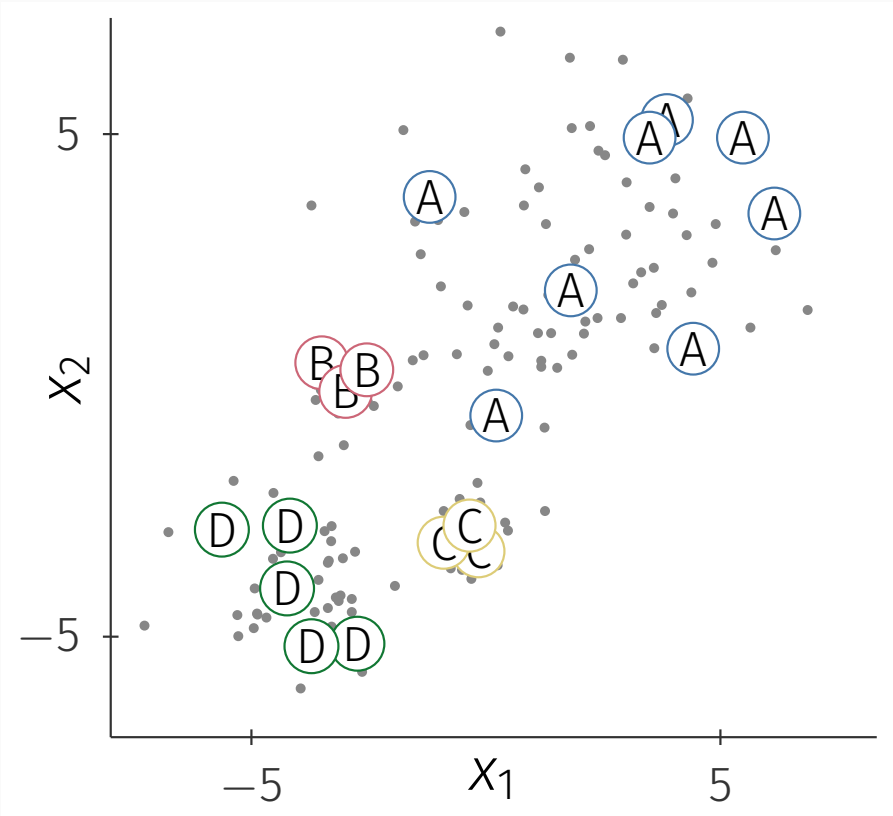
Consider the dataset below
In order to predict the classes of the unlabelled instances, you recommend to use

an EM-algorithm,

a transductive SVM algorithm,

a random forest, or

a graph-based approach



# Q2.3: Algorithm(s) of choice (ii)

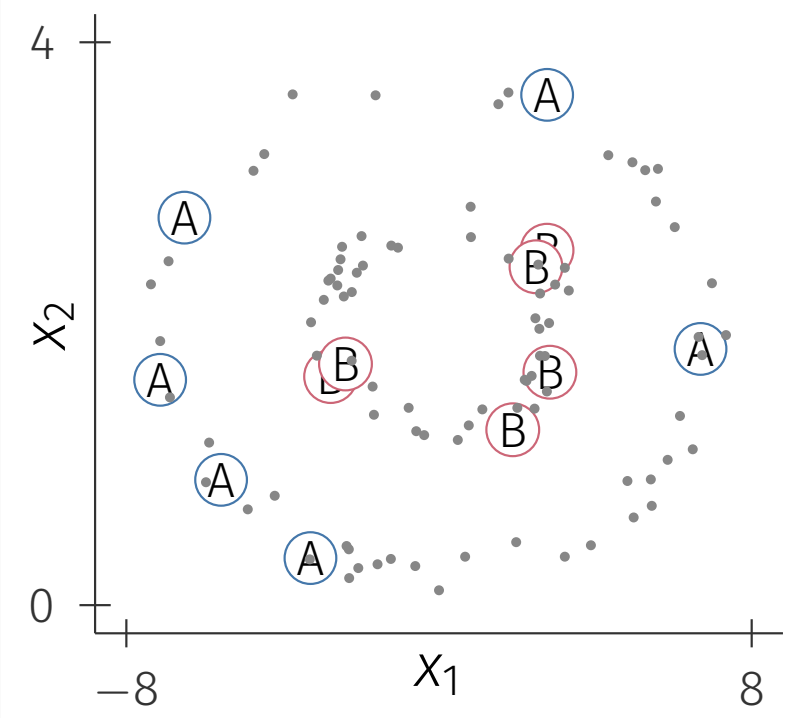
Consider the dataset below
In order to predict the classes of the unlabelled instances, you recommend to use

an EM-algorithm,

a transductive SVM algorithm,

a random forest, or

a graph-based approach



# Q2.4: Active learning strategy

In active learning, the uncertainty sampling strategy selects as query an instance with $\left\{ \begin{array}{c} \text{minimum} \\ \text{maximum} \end{array} \right\}$ label uncertainty.