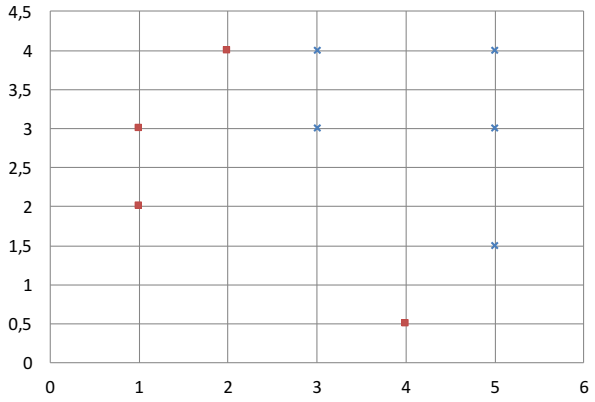
Theoretical task 12.

Recommendations: all solutions should be short, mathematically strict (unless qualitative explanation is needed), precise with respect to the stated question and clearly written.

1. Assume that the AdaBoost classification algorithm is trained on the dataset below. The weak classifiers set includes decision rules of the form $f(x_1, x_2) = sign\{x_i < \theta\}, i = 1, 2$.



   (a) Explain, what weak classifier will be obtained after the first iteration.

   (b) What object will have the largest weight after the first iteration? What is this weight value?

   (c) Explain, what weak classifier will be obtained after the second iteration.

   (d) Is it possible to choose the linear coefficients for these two weak classifiers such that their linear combination would classify the dataset with no errors?

2. Assume that the AdaBoost algorithm produced a set of weak classifiers $f_1, ..., f_T$. The ensemble classifier is defined as

$$F(x, w) = w_1 f_1(x) + ... + w_T f_T(x)$$

   Let's fit the linear coefficients $w_1, ..., w_T$ by minimizing the AdaBoost (exponential) loss function with gradient descent. Will the classifier F, obtained with a such training procedure, be equivalent to the classifier obtained with the AdaBoost procedure? Explain your answer.

3. Assume that we train the gradient boosting with decision trees as weak classifiers and the loss function $L(y_i, \hat{y}_i) = exp(-y_i \hat{y}_i)$. What labels should we use for training of a tree on the particular iteration if the current sum of the trained trees produces a vector of predicted labels $\hat{y}$?