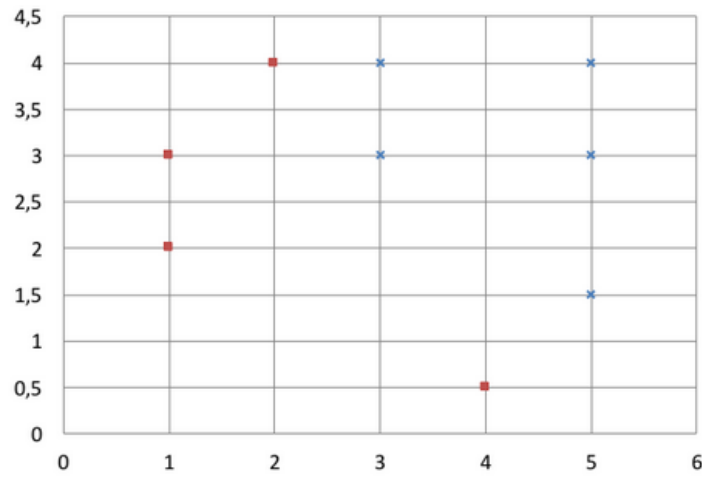


## Theoretical task 8.

*Recommendations: all solutions should be short, mathematically strict (unless qualitative explanation is needed), precise with respect to the stated question and clearly written. Solutions may be submitted in any readable format, including images.*

Submission link: [here](#)

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



- (a) Explain, what weak classifier will be obtained after the first iteration.
  - (b) What object(s) 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 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  $\hat{y}$ ?
  3. Consider linear regression model with  $d$  features  $g(x_1, x_2, \dots, x_d)$  and partial dependency plot of feature  $x_i$ :  $\phi(x_i)$ . Show that partial dependency plot for linear regression provides visualization of weight for feature  $x_i$ .