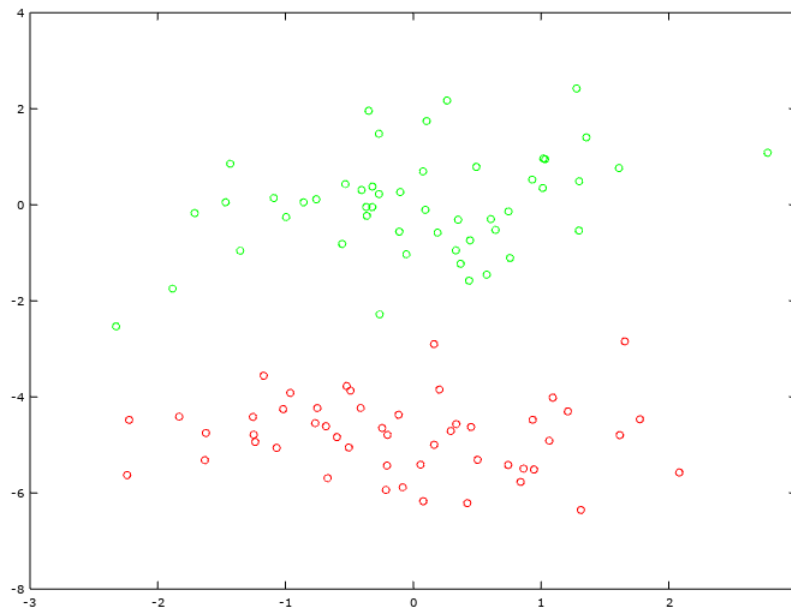


## EXERCISE 2

PROF. DR. VÍCTOR UC CETINA

### 1. BINARY CLASSIFICATION THROUGH LOGISTIC REGRESSION



- (1) Download the data file “data.txt” which contains 3 columns and 100 rows. The first 50 rows are positive examples (label 1) of points in 2 dimensions. The last 50 rows are negative examples (label 0) also in 2 dimensions. The first 5 rows of the file contains the following values (with precision 2):  
1.3 -0.54 1  
-2.3 -2.5 1  
-0.37 -0.047 1  
0.49 0.79 1  
1 0.95 1  
where the first two columns correspond to points in 2 dimensions, and the last column is the corresponding label.
- (2) Implement using python the Logistic Regression algorithm, so that you classify correctly both types of data.
- (3) Initialize the parameters of your model with random values in the interval  $(-0.01, 0.01)$ .

- (4) Plot the data points using one color for each class of data. Also, plot the classifier line that you found using logistic regression.
- (5) Prepare a report containing your final model (including parameters), your final  $\alpha$  value, and your graph.
- (6) Provide your code in the same document or as a link to your notebook in Google Colab (make sure to give sharing permissions to your notebook).

## 2. DEADLINE

- April 29th, 2022.

## 3. HINTS

- Use a linear model of the form  $z = \theta_0 + \theta_1 x_1 + \theta_2 x_2$ . Don't use a polynomial!
- This model should be evaluated in the sigmoid function  $g(\mathbf{x}) = \frac{1}{1+e^{-z}}$ .
- Once your algorithm has estimated the correct vector of parameters  $\theta$ , you can plot such model (the black line) using the function  $x_2 = (\theta_0 + \theta_1 x_1)(-1/\theta_2)$ .
- The solution should look like the following plot, where the blue line is the initial model (with random parameters) and the black line is the final answer, after 100 iterations.

