**Intro to ml - hw 4**

**Id: 208935916**

**Theoretical part:**

**1.**

We notice that the data is “linearly separable”, therefore there exists such that for every . Notice that for every the classifier is the same classifier as because for every .

By the correctness of the true classifier it holds that for every . For every define and by the former statement for all . Therefore,

Denote . So, by taking we get that

**2.**

Denote: and recall the following identity: . Let be the input of the neuron natural network. Consider the following learned parameter matrix and learned parameter vector :

Therefore, by using one hidden layer with 4 neurons and ReLU activations we have:

3.

1. Notice that the objective that we want to minimize contains each in squared form so the non-negativity of does not change the objective function value and so the optimal value. Moreover, let's say that we don't have the non-negativity constraints. Those, if there exists in the optimal solution which satisfies the constraint , then also satisfies the constraint and leads to same contribute in the objective function. Therefore, each optimal solution of the problem without the non-negativity constraints can be achieved with the non-negativity constraints. Oncourse all possible solutions of the problem with the non-negativity constraint contains in the possible solutions without these constraints.
2. The Lagrangian of this problem:
3. The derivative of the Lagrangian:

So,

Where is the j'th coordinate in the data vector. And we can present it as vectors:

Where is the matrix of all data set vectors and is the vector of all correspondent classifications.

We want to minimize the Lagrangian so set :

Now, by substituting the constraints we have:

1. The dual function is defined as:

The dual problem under the constraint is:

4.

First, we simplify the expression of :

(\*) Notice that for one-hot vector it holds that

Second, consider the partial derivative:

Therefore,

**Programming Assignment**