

Theoretical task 6.

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

1. In order to train a linear classification algorithm on linearly separable dataset, a student has implemented an SGD optimization method for perceptron loss $L(M) = [-M]_+$. Find a mistake in the following code listing:

```
Initialize weights:  $w = (w_1, \dots, w_d) = 0$ 
Until no mistakes on train set:
    i = GetRandomIndex()
    if  $y_i \langle x_i, w \rangle < 0$ :
         $w = w + \eta y_i x_i$ 
```

2. Assume we train a logistic regression model without regularization for some binary task. Does standardization of features of objects before training change a resulting classification model? Here we consider that two classification models are different iff there is an object for which they predict different labels. In addition, assume that we optimize the loss function perfectly and find a global optimum.

What about the case of logistic regression model with L2 regularization?

3. Assume we train a logistic regression model without regularization for some binary task. To optimize a loss function we use a GD optimization method. Does standardization of features of objects before training affect stability and convergence rate of optimization process? Think about the differences in gradient components along different dimensions in case of features with very different magnitudes.