Theoretical task 4.

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.

1. *The more parameters are included in machine learning algorithm, the more it is biased to overfitting. Indeed, overfitting means "flexibility"of the model towards each observation, that in turn means high "degree of freedom"(large number of parameters).*

   Consider classification results of two methods: Linear classifier (Figure 1) and K-Nearest Neighbour classifier (Figure 2). In m-dimensional space linear classifiers have about m weight parameters, while kNN has a single one – the number of nearest neighbours.

   Not only from that particular case it is clear, that despite having only one parameter, the decision boundary of kNN is more complex and flexible, as opposite to linear classifier solution. But that contradicts the valid argument about flexibility and the number of parameters! Why is this happening with kNN? Justify your answer.

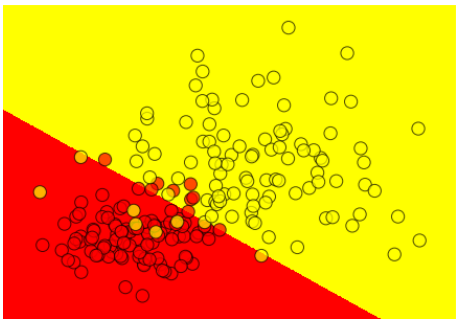

Fig. 1: Decision boundary of the linear classifier

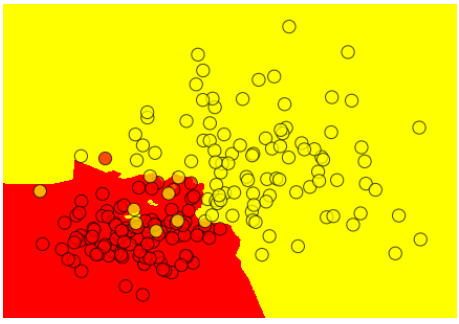

Fig. 2: Decision boundary of kNN

2. *A student has implemented perceptron algorithm for linearly separable dataset (it is possible to perfectly separate one class from another with a hyperplane). Find a mistake in he following code listing:*

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

3. *Considering logistic regression for binary classification*

   (a) *Show that if you multiply coefficients w and w_0 by a constant $c > 0$ you will end up with the same hyperplane*

   (b) *Consider a linearly separable dataset. Write down ML loss function for logistic regression (cross-entropy version). Show that the maximum likelihood solution for the logistic regression model is obtained with coefficients w and w_0 going to infinity*

   (c) *Name a technique that can help to overcome this issue*