

Theoretical task 5.

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

1. The more parameters are employed by machine learning algorithm, the more it has a tendency to overfit. 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.

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.

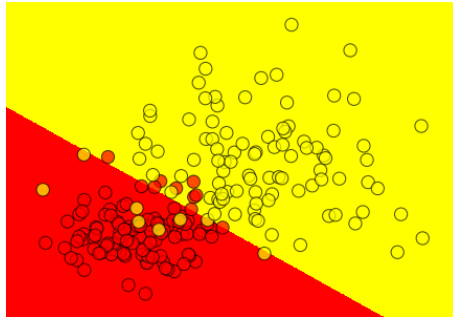


Fig. 1: Decision boundary of the linear classifier

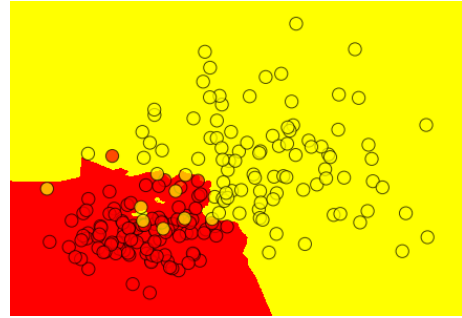


Fig. 2: Decision boundary of kNN

2. Prove that decision regions of linear classifier for multiclass classification task are convex (see slide 9 of the presentation "Linear methods of classification"). Decision region is a part of space which corresponds to one class (for all the points from this region the same class label is predicted). A definition of convex region can be found [here](#).
3. Show that for a linearly separable dataset, the maximum likelihood solution for the logistic regression model is obtained by finding a vector w , whose decision boundary $w^\top x = 0$ separates the classes and then taking the magnitude of w to infinity.

Name a commonly known technique that can help to overcome this issue.