Theoretical task 1.

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. *Consider classification with* 1*-nearest neighbour using euclidean distance:*
   (a) *Prove that the decision boundary for two training objects of different classes is linear.*
   (b) *Explain why decision boundaries separating classes in case of* 1*-nearest neighbour classifier are piecewise linear for* N *training objects and* C *classes.*
2. *Consider a training set of* N *objects with* D *features. Assume that each object has only* $s < D$ *nonzero features. Find computational complexity of classification of a new object with* 1*-nearest neighbour classifier with euclidean distance. (Remark: different objects may have different nonzero features but we explicitly know which ones)*
3. *Consider objects with categorical features. The simplest similarity measure for two objects with such features is overlap measure. It counts the number of features that match in both objects. The range of per-features similarity for the overlap measure is* $[0, 1]$*, with a value of* 0 *when there is no match, and a value of* 1 *when the feature values match.*

   Let's say that feature f *takes* P *possible values and some of them are more frequently occurring in the data than the others. For example, if* f *is a city of residence then Moscow is much more frequent value of* f *than Bobrov. Modify the overlap similarity measure in such way that it uses the information about frequency differences.*