Theoretical task 2.

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. *Prove that the complexity (number of elementary mathematical operations such as $+, -, *, /$ for scalars and boolean condition checks) for binary decision tree training from training set with N objects having D features each does not exceed $O(DN^2 \log(N))$. At this task suppose that the decision tree is balanced and each terminal node contains only one object.*

2. *Assume we use a particular feature i for splitting a certain internal node containing N objects. We want to determine the optimal threshold value h such that the sum of entropies for the children nodes $\{x|x_i < h\}$ and $\{x|x_i > h\}$ is the smallest. Propose the algorithm that founds the optimal threshold with complexity $O(Nlog(N))$.*

3. *Consider multiclass classification task with K classes. Objects in this task have L binary features. How many different decision trees of depth L can be constructed for this task? Two decision trees are different if there is an object in $\{0,1\}^L$ which they classify differently.*

4. *Suppose when building a decision tree, N objects $x_1, x_2, \ldots, x_N$ with labels $y_1, y_2, \ldots y_N$ appear in one leaf. Leaf predict constant value $\tilde{y}$. Show and proof which value $\tilde{y}$ must be choosen to optimize this error functions.*

   a) MSE (Mean Squared Error) for regression task

$$L = \frac{1}{N}\sum_{i=1}^{N}(y_i - \tilde{y})^2$$

   b) MAE (Mean Absolute Error) for regression task

$$L = \frac{1}{N}\sum_{i=1}^{N}|y_i - \tilde{y}|$$

   c) Log-Loss for classification task

$$L = -\frac{1}{N}\left(y_i \log(\tilde{y}) + (1 - y_i)\log(1 - \tilde{y})\right), \; y_i \in \{0,1\}, \; \tilde{y} \in [0;1]$$