Theoretical task 3

due January 24.

Remark: all solutions should be short, mathematically precise and contain proof unless qualitative explanation / intuition is needed. Solutions should be sent electronically to *ml.tasks@yandex.ru* and can be written in any clear and understandable format - latex, handwritten/scanned or other. Late submissions (by no more than 3 days) will be penalized by 50%, identical solutions will not be graded. The title of your e-mail should be "ICL homework <homework number> - <your first name and last name> ".

1. Suppose we know true class probabilities inside the node of the tree $p_1 = p(y = 1|x \in node), ...p_C = p(y = C|x \in node)$. Which of the following 2 strategies will always give lower expected error rate:

   (a) predict most common class $\widehat{y} = \arg\max_y[p_y]$

   (b) predict class randomly with probabilities proportional to class probabilities:

   $$\widehat{y} = \begin{cases} 1, & \text{with probability } p_1 \\ 2, & \text{with probability } p_2 \\ ... & ... \\ C & \text{with probability } p_C \end{cases}$$

   Prove your guess.

2. Consider binary classification performed with M classifiers $f_1(x), ...f_M(x)$. Let probability of mistake be constant $p \in (0, \frac{1}{2})$: $p(f_m(x) \neq y) = p\,\forall m$ and suppose all models make mistakes or correct guesses independently of each other. Let $F(x)$ be majority voting combiner. Prove that $\forall (x, y)\, p(F(x) \neq y) \to 0$ as $M \to \infty$.

   Hint: consider $\xi_m = \mathbb{I}[\text{model } m \text{ made mistake}]$. Look at limit properties of $\sum_{m=1}^{M} \xi_m / M$.

3. Write out the targets z_i, $i = 1, 2, ...N$ for gradient boosting for each of the following losses:

   (a) $[-f(x)y]_+ = \max\{0, -f(x)y\}$, $y \in \{+1, -1\}$.

   (b) $e^{-f(x)y}$, $y \in \{+1, -1\}$.

   Besides each target provide intuition, how it changes for well/poorly classified objects.