Машинное обучение, ФКН ВШЭ
Теоретическое домашнее задание №8

Задача 1. Покажите, что для выпуклой функции потерь $L(y, z)$ ошибка композиции $a(x) = \frac{1}{N}\sum_{n=1}^{N} b_n(x)$ на некоторой выборке $X = \{(x_i, y_i)\}_{i=1}^{\ell}$ не превзойдет среднего по всем базовым алгоритмам b_n значения ошибки на этой же выборке.

Задача 2. Напомним, что критерий информативности в вершине R решающего дерева в общем случае определяется следующим образом:

$$H(R) = \min_{c \in \mathbb{Y}} \frac{1}{|R|} \sum_{(x_i, y_i) \in R} L(y_i, c).$$

Предложите функцию потерь $L(y, z)$, при использовании которой данный критерий совпадёт с критерием информативности, используемом при обучении решающих деревьев в XGBoost.

Задача 3. Напомним, что при использовании градиентного бустинга над решающими деревьями при добавлении N-ого базового алгоритма с J листьями решается следующая оптимизационная задача:

$$\sum_{i=1}^{\ell} L(y_i, a_{N-1}(x_i) + \sum_{j=1}^{J} \gamma_{Nj}[x_i \in R_j]) \to \min_{\{\gamma_{Nj}\}_{j=1}^{J}},$$

где γ_{Nj} — прогноз N-ого базового алгоритма в j-ом листе, $R_j \subset \mathbb{X}$ — подмножество пространства объектов, отвечающее j-ому листу.

Найдите оптимальные значения $\{\gamma_{Nj}\}_{j=1}^{J}$ для следующих функций потерь:

1. $L(y, z) = (y - z)^2$;
2. $L(y, z) = |y - z|$;
3. $L(y, z) = e^{-yz}$ (для $y \in \{-1, +1\}$).

Задача 4. Рассмотрим задачу бинарной классификации, $\mathbb{Y} = \{0, 1\}$. Будем считать, что все алгоритмы из базового семейства $\mathcal{B}$ возвращают ответы из отрезка $[0; 1]$, которые можно интерпретировать как вероятности принадлежности объекта к классу +1. Выпишите формулы для поиска базового алгоритма b_n и коэффициента γ_n в градиентном бустинге при использовании отрицательного логарифма правдоподобия (negative log-likelihood) в качестве функции потерь:

$$L(y, z) = -y \log z - (1 - y) \log(1 - z).$$