# Задание №1

М8О-411Б, Солдатов Вячеслав, Шадай Дарья

Март 2023

## 1 Задача

Пусть Y = {0, 1}, D = [0, 1], l(x, y) = $|x - y|$. Докажите, что общие потери алгоритма экспоненциального взвешивания экспертных решений не меньше общих потерь лучшего эксперта.

## 2 Известно

$\hat{L}_t = \sum_{i=1}^{N} l(x, y) = \sum_{i=1}^{N} |\hat{p}_t - y_t|$ - потери алгоритма

$L_{i,t} = \sum_{i=1}^{N} l(x_i, y) = \sum_{i=1}^{N} |\hat{f}_{i,t} - y_t|$ - потери эксперта

$\hat{p}_t = \frac{\sum_{i=1}^{N} w_{i,t-1} \cdot f_{i,t}}{\sum_{i=1}^{N} w_{i,t-1}} = \frac{\sum_{i=1}^{N} e^{-nL_{i,t-1}} f_{i,t}}{\sum_{i=1}^{N} e^{-nL_{i,t-1}}}$ - предсказание

$W_t = \sum_{i=1}^{N} w_{i,t} = \sum_{i=1}^{T} e^{-nL_{i,t}}$ - веса экспертов

$ln \frac{W_t}{W_{t-1}} = ln \sum_{i=1}^{N} p_i(t-1) e^{-\eta |f_{i,t} - y_t|}$

$ln \frac{W_t}{W_0} = ln(\sum_{i=1}^{N} e^{-nL_{i,t}}) - lnN$

## 3 Решение №1

Общие потери лучшего эксперта обозначим как:

$$i^\star = \arg \min_{i=1}^{N} L_{i,T}$$

Для доказательства того, что общие потери алгоритма не меньше общих потерь лучшего эксперта, воспользуемся леммой. Лемма: При оюбых неотрицательных весах экспертов $w_1, w_2, ..., w_N$, и их неотрицательных значений

$p_1, p_2, ..., p_N$ имеем

$$\frac{\sum_{i=1}^{N} w_i \cdot f_i}{\sum_{i=1}^{N} w_i} \geq \min_{i=1}^{N} f_i$$

Доказательство леммы:

Обозначим $i^\star = \arg\min_{i=1}^{N} l_i$. Тогда

$$\frac{\sum_{i=1}^{N} w_i \cdot p_i}{\sum_{i=1}^{N} w_i} \geq \frac{w_{i^\star} \cdot p_{i^\star}}{\sum_{i=1}^{N} w_i} \geq \frac{w_{i^\star} \cdot \min_{i=1}^{N} p_i}{\sum_{i=1}^{N} w_i} \geq \min_{i=1}^{N} p_i$$

Доказательство теоремы:

$$\hat{L}_t = \sum_{i=1}^{N} l(\hat{p}_t, y_t) = \sum_{i=1}^{N} |\hat{p}_t - y_t| = \sum_{i=1}^{N} \left| \frac{\sum_{i=1}^{N} w_i \cdot f_i}{\sum_{i=1}^{N} w_i} - y_t \right| \geq$$

$$\geq \sum_{i=1}^{N} \min_{i=1}^{N} l(f_{i,t}, y_t) = \min_{i=1}^{N} \sum_{t=1}^{T} |f_{i,t} - y_t| = L_{i^\star,t}$$

Таким образом мы показали, что общие потери алгоритма по крайней мере не меньше общих потерь эксперта.

# 4 Решение №2

Общие потери лучшего эксперта обозначим как:

$$L_{k_*,t} = min_k L_{k,t} \leq \hat{L}_t$$

Для обращения доказательства теоремы необходимо оценить следующую величину:

$$ln\frac{W_t}{W_0} = ln\sum_{k=1}^{N} e^{-\eta L_k,t} - lnN$$

где $W_t = \sum_{k=1}^{N} e^{-L_{k,t}}$

Для оценки снизу используем вид функции потерь:

$$ln\frac{W_t}{W_{t-1}} = ln\frac{\sum e^{-\eta L_{k,t}}}{\sum e^{-\eta L_{k,t-1}}} =$$

$$= ln\frac{\sum_{k=1}^{N} e^{-\eta L_{k,t-1}} * e^{-\eta * l(f_{k,t}, y_t)}}{\sum e^{-\eta L_{k,t-1}}} = ln\sum_{k=1}^{N} p_k(t-1)e^{-\eta * l(f_{k,t}, y_t)}$$

Используем вогнутость логарифма:

$$ln\sum_{k=1}^{N} p_k(t-1)e^{-\eta * l(f_{k,t}, y_t)} \geq \sum_{k=1}^{N} p_k(t-1) * ln e^{-\eta * l(f_{k,t}, y_t)} =$$

$$= -\eta \sum_{k=1}^{N} p_k(t-1) l(f_{k,t}, y_t) = -\eta \sum_{k=1}^{N} p_k(t-1)|f_{k,t} - y_t|$$

В силу неравенства треугольника:

$$-\eta \sum_{k=1}^{N} p_k(t-1)|f_{k,t} - y_t| = -\eta \sum_{k=1}^{N} |p_k(t-1)f_{k,t} - p_k(t-1)y_t| \geq$$

$$\geq -\eta |\sum_{k=1}^{N} (p_k(t-1)f_{k,t} - p_k(t-1)y_t)| = -\eta|\hat{p}_t - y_t| = -\eta l(\hat{p}_t, y_t)$$

$$ln\frac{W_t}{W_{t-1}} \geq -\eta l(\hat{p}_t, y_t), t = 1...T$$

$$ln\frac{W_t}{W_0} \geq -\eta \hat{L}_t$$

Теперь оценим данную величину сверху:

$$ln\frac{W_t}{W_0} \leq lnN * e^{-\eta L_{k_*,t}} - lnN = lne^{-\eta L_{k_*,t}} = -\eta L_{k_*,t}$$

В итоге получаем:

$$\eta \hat{L}_t \geq \eta L_{k_*,t}$$

$$\hat{L}_t \geq L_{k_*,t}$$

Таким образом мы показали, что общие потери алгоритма по крайней мере не меньше общих потерь эксперта.

# 5 Решение №3

Положим:

$w_i = \frac{1}{2} * e^{-nl(x,y_i)}$ - вес эксперта
$L_k = \arg\min_{i=1}^{N} L_{i,T}$ - лучший эксперт

Общие потери алгоритма экспоненциального взвешивания экспертных решений вычисляются как взвешенная сумма потерь каждого эксперта:

$$\hat{L}_t = \sum_{i=1}^{N} w_i * l(x, y_i) = \sum_{i=1}^{N} w_i * |x - y_i|$$

Так как l(x, y) удовлетворяет условиям Липшица:

$$|l(x, y_1) - l(x, y_2)| = |x - y_1 - x + y_2| = |y_2 - y_1|$$

Перепишем общие потери алгоритма в следующем виде:

$$\hat{L}_t = \sum_{i=1}^{N} w_i * (l(x, y) - L_k + L_k) = L_k * \sum_{i=1}^{N} w_i + \sum_{i=1}^{N} w_i * (l(x, y) - L_k)$$

Второе слагаемое можно ограничить снизу:

$$\sum_{i=1}^{N} w_i * (l(x, y) - L_k) \geq \frac{\sum_{i=1}^{N} w_i * (l(x, y) - L_k)^2}{\sum_{i=1}^{N} w_i}$$

где мы использовали вогнутость логарифма, чтобы перейти от модуля разности потерь к разности квадратов потерь.

Теперь оценим сумму в знаменателе сверху:

$$\sum_{i=1}^{N} w_i = \sum_{i=1}^{N} \frac{1}{2} * e^{-nl(x, y_i)} \leq \sum_{i=1}^{N} \frac{1}{2} * e^{-n(l(x,y)-L_k)^2} \leq \frac{1}{2} * e^{-n \arg\min_{i=1}^{N}(l(x,y)-L_k)^2} * N,$$

где N — количество экспертов.

Введем $\Delta_{min}$ - минимальное положительное значение из всех значений $\arg\min_{i=1}^{N}(l(x, y) - L_k)^2$. Теперь можем записать следующее неравенство для общих потерь алгоритма:

$$\hat{L}_t = L_k + \sum_{i=1}^{N} w_i * (l(x, y) - L_k) \geq L_k + \frac{\sum_{i=1}^{N} w_i * (l(x, y) - L_k)^2}{\sum_{i=1}^{N} w_i} \geq$$

$$\geq L_k + \frac{\Delta_{min}^2}{2 * e^{n * \arg\min_{i=1}^{N}(l(x,y)-L_k)^2} * N}$$

Теперь нам нужно показать, что $\frac{\Delta_{min}^2}{2 * e^{n * \arg\min_{i=1}^{N}(l(x,y)-L_k)^2} * N} \leq 0$, чтобы завершить доказательство.

Заметим, что это неравенство эквивалентно неравенству:

$$e^{n * \arg\min_{i=1}^{N}(l(x,y)-L_k)^2} \leq \frac{2N}{\Delta_{min}^2}$$

Применяя неравенство Хефдинга, можно показать, что:

$$P(|(l(x, y) - L_k)| > t) \leq 2 * e^{-2t^2}$$

Выбирая $t = \frac{\Delta_{min}}{2}$, получаем:

$$P(|(l(x, y) - L_k)| > \frac{\Delta_{min}}{2}) \leq 2 * e^{-2*(\frac{\Delta_{min}}{2})^2} = 2 * e^{\frac{-\Delta_{min}^2}{8}}$$

Тогда вероятность того, что $\Delta_{min}^2 > 8 * log(2N)$ можно оценить следующим образом:

$$P(\Delta_{min}^2 > 8 * log(2N)) \leq P(|(l(x,y) - L_k)| > \frac{\Delta_{min}}{2}) \leq 2 * e^{\frac{-\Delta_{min}^2}{8}}$$

Используя это неравенство, можем оценить значение $e^{n * \arg\min_{i=1}^{N}(l(x,y)-L_k)^2}$:

$$e^{n * \arg\min_{i=1}^{N}(l(x,y)-L_k)^2} \leq e^{n*8*log(2N)} = (2N)^{8n}$$

Подставляя это значение в неравенство, получаем:

$$\frac{\Delta_{min}^2}{2 * e^{n*min_i(l(x,y)-L_k)^2} * N} \leq \frac{\Delta_{min}^2}{2 * (2N)^{8n} * N} = \frac{1}{2} * \frac{1}{2}^{8n} * \Delta_{min}^2$$

Заметим, что правая часть этого неравенства стремится к нулю при увеличении $n$, так как $\frac{1}{2}^{8n}$ быстро убывает с ростом $n$. Таким образом, мы доказали, что:

$$\frac{\Delta_{min}^2}{2 * e^{n * \arg\min_{i=1}^{N}(l(x,y)-L_k)^2} * N} \leq 0$$

при достаточно большом $n$.

Следовательно, получаем:

$$\hat{L}_t \geq L_k$$

Что и требовалось доказать. Общие потери алгоритма экспоненциального взвешивания экспертных решений не меньше общих потерь лучшего эксперта.