# Домашнее задание №6 по курсу "Машинное обучение"

Прокопенко Тимофей

## Задача 1.

1. Постройте пример того, что 0-1 функция потерь может обладать локальным минимумом, не являющимся глобальным.

   Другими словами, приведите пример выборки $S \in (X \times \{\pm 1\})^m$ (например, для пространства примеров $X = \mathbb{R}^2$), для которой существует вектор $w$ и некоторый $\epsilon > 0$, что выполняется

   1. Для любого $w'$, такого, что $\|w - w'\| \leq \epsilon$, выполняется $L_S(w) \leq L_S(w')$. Это означает, что $w$ — локальный минимум $L$, где $L$ — 0-1 функция потерь.
   2. Существует некоторый $w^*$, такой, что $L_S(w^*) < L_S(w)$. Это означает, что $w$ не является глобальным минимумом.

Решение: Возьмем выборку следующего вида

$S_1 = (1, 0) \quad y_1 = 1$

$S_2 = (0.5, 1) \quad y_2 = -1 \qquad n = 2$

Выберем $w_0 = (0, 1)$, $\varepsilon = 0,2$.

В качестве $L_S$ рассмотрим

$$L_S(w) = \sum_{i=1}^{n} \left( \text{sign}(\langle S_i, w \rangle) \mathrel{!=} y_i \right), \quad \text{true} = 1, \ \text{false} = 0$$

причем пусть $\text{sign}(0) = 1$, т.к. мы решаем задачу $\pm 1$ классификации

1) Теперь докажем, что для $\forall w'$ такого, что $\| w_0 - w' \| \le \varepsilon$ выполняется $L_s(w_0^*) \le L_s(w')$. В качестве нормы берем $l$-норму.

$$L_s(w_0) = (sign(\langle (1,0), (0,1) \rangle) != 1) + (sign(\langle (0.5,1), (0,1) \rangle) != -1) =$$

$$= (sign(0) != 1) + (sign(1) != -1) = 0 + 1 = 1$$

Рассмотрим крайние значения $\varepsilon$-интервала:

$$L_s(w_0 + \tfrac{\varepsilon}{2}) = (sign(\langle (1,0), (0.1, 1.1) \rangle) != 1) + (sign(\langle (0.5,1), (0.1, 1.1) \rangle) != -1)$$

$$= (sign(0,1) != 1) + (sign(1.15) != -1) = 0 + 1 = L_s(w_0)$$

$$L_s(w_0 - \tfrac{\varepsilon}{2}) = (sign(\langle (1,0), (-0.1, 0.9) \rangle) != 1) + (sign(\langle (0.5,1), (-0.1, 0.9) \rangle) != -1) =$$

$$= (sign(-0,1) != 1) + (sign(0,85) != -1) = 1 + 1 = 2 \ge L_s(w_0)$$

Очевидно, что внутри промежутка $L_S(w')$ не будет принимать значение ноль. Таким образом, $w_0$ – локальный минимум $L_S$.

2) Возьмем $w^* = (1, -1)$.

$L_S(w^*) = (\text{sign}(<(1, 0), (1, -1)>) != 1) + (\text{sign}(<(0{,}5, 1)(1, -1)>) != -1) = (\text{sign}(1) != 1) + (\text{sign}(-0{,}5) != -1) = 0 + 0 = 0 < L_S(w_0)$

Значит $w_0$ не является глобальным минимумом.

## Задача 2.

2. Рассмотрим задачу обучения логистической регрессии. Пусть $H = X = \{x \in \mathbb{R}^d : \|x\| \le B\}$ для некоторой скалярной величины $B$, пусть $y = \pm 1$, и пусть функция потерь имеет вид

$$l(w, (x, y)) = \log(1 + \exp(-y\langle w, x\rangle))$$

Покажите, что задача одновременно и выпукло-липшицево-ограничена, и выпукло-гладко-ограничена. Приведите соответствующие параметры липшицевости и гладкости.

**Решение:**

Для того чтобы задача была выпукло-липшицево-ограничена с параметрами $\rho, B$ необходимо выполнение двух условий:

1) H - выпуклое множество и $\forall w \in H$ ||w||$\leq$B

2) $\forall z \in Z$ l(w,z) - выпуклая и $\rho$-липшицева.

Рассмотрим 1 условие. Ограниченность нормы вектора обеспечивается самой формой задания множества H, как и выпуклость данного множества.

Рассмотрим 2 условие. Функция l(w, (x, y)) является выпуклой по лемме о композиции выпуклой и линейной функции, где выпуклая функция g(a)=log(1+exp(a)) будет использоваться нами при некоторых последующих доказательствах.

$\rho$-липшицевость будем доказывать с помощью первой производной:

$$|g'(a)| = |\frac{exp(a)}{1 + exp(a)}| = |\frac{1}{1 + exp(-a)}| \le 1$$

Следовательно, функция g(a) 1-липшицева. Применим лемму о композиции липшицевых функций, учитывая что линейная функция $\rho$-липшицева с параметром $\rho$=||x||(доказывается с помощью определения липшицевости и неравенства Коши-Шварца), и получим, что l(w,(x,y)) - $\rho$-липшицева с параметром $\rho$=||x||*1.

Перейдем к доказательству выпукло-гладко-ограниченности с параметрами $\beta$, B. Для этого нужно, чтобы выполнялись 2 условия:

1) H - выпуклое множество и $\forall w \in H$ ||w||$\leq$B

2) $\forall z \in Z$ l(w,z) - выпуклая, неотрицательная и $\beta$-гладкая.

Выполнение первого условия и выпуклость функции l(w, (x,y)) мы показали раньше. Очевидно, что функция логарифма неотрицательна при любых параметрах. Таким образом, осталось доказать $\beta$-гладкость. Для начала рассмотрим вторую производную g(a):

$$|g''(a)| = |\frac{exp(-a)}{(1 + exp(-a))^2}| = |\frac{1}{(1 + exp(a))(1 + exp(-a))}| \le \frac{1}{4}$$

Таким образом, функция g(a) - $\frac{1}{4}$-гладкая. Воспользуемся леммой о композиции гладкой и линейной функции, получим, что функция l(w, (x,y))
$\frac{||x||^2}{4}$-гладкая.

Таким образом, задача логистической регрессии одновременно и липшицево-выпукло-ограничена, и выпукло-гладко-ограничена. Учтем, что ||x||$\leq$B по условию и, таким образом, параметр липшицевости - B, гладкости - $\frac{B^2}{4}$.

## Задача 3.

3. Рассмотрим задачу полупространств с hinge-loss. Пусть множество примеров — евклидов шар радиуса $R$, то есть $X = \{x : \|x\|_2 \leq R\}$. Множество меток $Y = \{\pm 1\}$. Функция потерь $l(w,(x,y)) = \max\{0, 1 - y\langle w, x\rangle\}$. Данная функция потерь является выпуклой.

   Покажите, что эта функция потерь также является $R$-липшицевой.

**Решение:**

Решение: Найдем $l'(w,(x,y))$:

$$l'(w,(x,y)) = \begin{cases} (0,0\ldots 0) & \text{при } y\langle w,x\rangle \geq 1 \\ (-yx_1,-yx_2,\ldots,-yx_n) & \text{при } y\langle w,x\rangle < 1 \end{cases}$$

В первом случае $|l'(w,(x,y))| = 0$, во втором

$$|l'(w,(x,y))| = \sqrt{\sum_{i=1}^{n} y^2 x_i^2} = \underbrace{\|x\| \leq R}_{\text{по условию}}$$

Таким образом, $|l'(w,(x,y))| \leq R$, т.е. данная функция потерь является $R$-липшицевой.