# Problem set 1

Красильников Иван

22 сентября 2010 г.

## Задача 1

Да. $a(x) = \text{sign}\left(-3 + \text{sign}(x_1 > 0) + \text{sign}(x_1 < 1) + \text{sign}(x_2 > 0) + \text{sign}(x_2 < 1)\right)$.

## Задача 2

Нет. Нельзя, например, описать $\{(x, y) \,|\, xy > 0\}$ (т.е. решить XOR), что можно показать невозможностью отделить точки $(1, 1), (-1, -1)$ от $(-1, 1), (1, -1)$ решающими правилами вида $\text{sign}(f(x) + g(y))$:

$$\begin{cases} f(1) + g(1) > 0, \\ f(-1) + g(-1) > 0, \\ f(-1) + g(1) < 0, \\ f(1) + g(-1) < 0, \end{cases} \implies \begin{cases} f(1) > -g(1) \\ f(-1) > -g(-1) \\ f(-1) < -g(1), \\ f(1) < -g(-1), \end{cases} \implies$$

$f(1) > -g(1) > f(-1) > -g(-1) > f(1)$, что неверно.

## Задача 3

**3.1.** Первый базовый классификатор: $h_1(x, y) = \text{sign}(x - 2.5)$. Он положителен при $x > 2.5$, отрицателен при $x < 2.5$.

Здесь должен быть рисунок, но у меня нет времени на его оформление.

Неправильно классифицировался один пример из 9, поэтому

$$\varepsilon_1 = \frac{1}{9}, \quad \alpha_1 = \frac{1}{2} \log \frac{1 - \varepsilon_1}{\varepsilon_1} = \log \sqrt{8} \approx 1.03972$$

**3.2.** Ненормализованные веса: у неправильно классифицированного объекта $(4, 0.5)$ вес $\frac{1}{9} e^{\alpha_1} = \frac{1}{9}\sqrt{8}$, у остальных веса $\frac{1}{9} e^{-\alpha_1} = \frac{1}{9}\frac{1}{\sqrt{8}}$. Сумма весов: $\frac{1}{9}\left(8\frac{1}{\sqrt{8}} + \sqrt{8}\right) = \frac{2}{9}\sqrt{8}$. Веса после нормализации: у $(4, 0.5)$ вес $\frac{1}{2}$, у остальных: $\frac{1}{16}$.

**3.3.** Равно сумме весов неправильно классифицированных объектов, то есть весу одного объекта $(4, 0.5)$: $\frac{1}{2}$.

**3.4.** Столь большой вес $(4, 0.5)$ накладывает ограничие на новый базовый классификатор – он должен обязательно правильно классифицировать этот объект, иначе его взвешенная ошибка не будет меньше $\frac{1}{2}$. Из таких классификаторов взвешенную ошибку минимизирует $\text{sign}(4.5 - x)$. Он положителен при $x < 4.5$ и отрицателен при $x > 4.5$.

**3.5.** Нет, так как при любых коэффициентах функция $\alpha_1 \cdot \text{sign}(x - 2.5) + \alpha_2 \cdot \text{sign}(4.5 - x)$ будет постоянной на интервале $(2.5, 4.5)$, а там находятся обучающие примеры разных классов.

## Задача 4

После обучения $t$-го базового классификатора $h_t$ с весами объектов $w_i^{(t)}$ (нормализованные так, чтобы их сумма была равна единице), веса в алгоритме AdaBoost пересчитываются по следующей формуле:

$$w_i^{(t+1)} = Z_t w_i^{(t)} e^{-y_i \alpha_t h_t(x_i)},$$

где $Z_t$ – множитель для нормализации новых весов:

$$\begin{aligned}\frac{1}{Z_t} &= \sum_i w_i^{(t)} e^{-y_i \alpha_t h_t(x_i)} = \sum_{i:y_i = h_t(x_i)} w_i^{(t)} e^{-\alpha_t} + \sum_{i:y_i \neq h_t(x_i)} w_i^{(t)} e^{\alpha_t} \\ &= (1-\varepsilon_t) e^{-\alpha_t} + \varepsilon_t e^{\alpha_t},\end{aligned}$$

а $\varepsilon_t$ – функционал взвешенной ошибки классификатора $h_t$ с весами $w_i^{(t)}$:

$$\varepsilon_t = \sum_{i:y_i \neq h_t(x_i)} w_i^{(t)}.$$

Значение функционала взвешенной ошибки для алгоритма $h_t$ с новыми, пересчитанными и нормализованными весами $w_i^{(t+1)}$:

$$\begin{aligned}\varepsilon_t' &= \sum_{i:y_i \neq h_t(x_i)} w_i^{(t+1)} = \sum_{i:y_i \neq h_t(x_i)} Z_t w_i^{(t)} e^{-y_i \alpha_t h_t(x_i)} \\ &= Z_t \sum_{i:y_i \neq h_t(x_i)} w_i^{(t)} e^{\alpha_t} = Z_t \varepsilon_t e^{\alpha_t} = \frac{\varepsilon_t e^{\alpha_t}}{(1-\varepsilon_t) e^{-\alpha_t} + \varepsilon_t e^{\alpha_t}} \\ &= \frac{1}{1 + \frac{1-\varepsilon_t}{\varepsilon_t} e^{-2\alpha_t}} = \frac{1}{1+1} = \frac{1}{2},\end{aligned}$$

так как

$$\alpha_t = \frac{1}{2} \log \frac{1-\varepsilon_t}{\varepsilon_t}, \quad e^{-2\alpha_t} = e^{-2 \cdot \frac{1}{2} \log \frac{1-\varepsilon_t}{\varepsilon_t}} = \frac{\varepsilon_t}{1-\varepsilon_t}.$$

**Ответ:** $\frac{1}{2}$.

## Задача 5

Пусть задана выборка $\{(x_i, y_i)\}_{i=1}^n$, $x_i \in \mathbb{R}^m$, $y_i \in \mathbb{R}$, и требуется найти $f(x)$, минимизирующую сумму квадратов ошибок: $\sum_{i=1}^n (f(x_i) - y_i)^2$.

На первом шаге положим $f_0(x) = \frac{1}{n}\sum_{i=1}^n y_i$ — наилучшее константное приближение.

Далее, следуя общей концепции бустинга, на $t$-м шаге будем искать $f_t(x)$ в виде $f_t(x) = f_{t-1}(x) + h(x)$, где $h(x)$ – функция, возвращаемая базовым алгоритмом обучения, критерий выбора которой – минимизация квадратичной ошибки функции $f_t(x)$:

$$\sum_{i=1}^n (y_i - f_t(x_i))^2 \to \min_h$$

$$\sum_{i=1}^n ([y_i - f_{t-1}(x_i)] - h(x_i))^2 \to \min_h$$

Таким образом, для выбора $h(x)$ необходимо решить задачу регрессии на выборке $\{(x_i, y_i - f_{t-1}(x_i))\}_{i=1}^n$. Это уже задача для базового алгоритма регрессии.

Алгоритм построен. ■

Псевдокод:

1. $f_0(x) := \frac{1}{n}\sum_{i=1}^n y_i$.
2. Для $t = 1, 2, \ldots, T$:
   (a) Построить функцию $h(x)$ при помощи базового алгоритма для решения задачи регрессии:
   $$h(x) = \arg\max_h \sum_{i=1}^n (h(x_i) - r_i)^2,$$
   где $r_i = y_i - f_{t-1}(x_i)$.
   (b) $f_t(x) := f_{t-1}(x) + h(x)$:
3. Вывести $f_T(x)$.