

Построение Adaboost

1. Загрузите выборку data_adaboost.csv.
2. В качестве базового алгоритма будем использовать решающий пень. Решающий пень – это дерево решений глубины 1 (одно разбиение).
3. В случае с вещественными признаками порог разбиения выбирается следующим образом: значения признаков упорядочиваются по возрастанию. Далее рассматриваются все случаи, где меняется значение целевой переменной. Порог – среднее между значениями тех признаков, с которыми меняется значение целевой переменной.

x	y	$thres$
-2	0	-1
0	1	
4	1	
4.5	1	5.1
5.7	0	
7	0	8.2
9.4	1	

4. Определите все возможные пороги разбиения по x_1 и x_2 для выборки data_adaboost.csv.
5. Задайте начальные значения весов объектов $w_i^{(0)} = \frac{1}{l}$, $i = 1, \dots, l$.
6. Следуя алгоритму, постройте композицию решающих пней. Количество базовых алгоритмов: $T = 3$.
7. Пользуясь полученной композицией, классифицируйте точку (4,4).

Алгоритм Adaboost

Для всех $t = 1, \dots, T$:

1. Обучить базовый алгоритм b_t : $b_t = \arg \min_j \epsilon_j$, где ϵ_j – ошибка алгоритма, вычисляемая по формуле:

$$\epsilon_j = \sum_{i=1}^l w_i [y_i \neq b_j(x_i)].$$

2. Вычислить $\alpha_t = \frac{1}{2} \log \frac{1-\epsilon_t}{\epsilon_t}$.
3. Обновить веса объектов: $w_i^{(t)} = w_i^{(t-1)} \exp(-\alpha_t y_i b_t(x_i))$, $i = 1, \dots, l$.
4. Нормировать веса объектов: $w_0^{(t)} = \sum_{i=1}^l w_i^{(t)}$, $w_i^{(t)} = \frac{w_i^{(t)}}{w_0^{(t)}}$, $i = 1, \dots, l$.
5. Вернуть $\sum_{t=1}^T \alpha_t b_t$.