Машинное обучение.
Домашнее задание №3

Задача 1. На семинаре был рассмотрен пример с бутстрэппингом, где строилась композиция из нескольких функций регрессии. Откажемся от предположения о несмещенности и некоррелированности ошибок.

Пользуясь неравенством Йенсена, покажите, что среднеквадратичная ошибка композиции не превосходит среднюю ошибку отдельных алгоритмов:

$$E_n \leqslant E_1.$$

Задача 2. *Решающим пнем* называется классификатор вида

$$a(x; j, t) = \begin{cases} +1, & x_j < t; \\ -1, & \text{иначе.} \end{cases}$$

Его параметрами являются номер признака j и порог t.

Рассмотрим одномерную выборку, состоящую из четырех объектов $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $x_4 = 4$ с ответами $y_1 = +1$, $y_2 = +1$, $y_3 = -1$, $y_4 = +1$.

Применим к этой выборке алгоритм AdaBoost с решающими пнями в качестве базовых алгоритмов.

1. Какими будут веса на первой итерации?
2. Какой базовый классификатор будет выбран на первой итерации?
3. Какой будет выбран вес γ_1?
4. Какими будут веса на второй итерации?

Задача 3. Пусть $X^\ell = \{x_1, \ldots, x_\ell\} \subset \mathbb{R}$ — произвольная одномерная обучающая выборка. Покажите, что при любых ответах $(y_i)_{i=1}^\ell$ на этих объектах существует композиция вида

$$a(x) = \operatorname{sign} \sum_{n=1}^{N} \gamma_n a_n(x)$$

над решающими пнями, не допускающая ошибок на обучающей выборке X^ℓ. Покажите, что в ней будет не более $2\ell + 2$ различных классификаторов (классификаторы считаются одинаковыми, если они дают одинаковые ответы на обучающей выборке).

Задача 4. Пусть $X^\ell = \{x_1, \ldots, x_\ell\} \subset \mathbb{R}^d$ — произвольная d-мерная обучающая выборка. Рассмотрим *радиальные* базовые функции:

$$a(x; i) = y_i \exp\left(-\beta \|x - x_i\|^2\right).$$

Параметром такой функции является индекс i объекта обучающей выборки; величина $\beta > 0$ является гиперпараметром и считается фиксированной в нашей задаче. Покажите, что при любых ответах $(y_i)_{i=1}^\ell$ существует взвешенная композиция радиальных функций

$$a(x) = \operatorname{sign} \sum_{n=1}^{N} \gamma_n a_n(x),$$

не допускающая ошибок на обучающей выборке X^ℓ.